

UNIVERSITÉ DU QUÉBEC À MONTRÉAL

MODÉLISATION DE LA VARIANCE DANS L'ANALYSE
STOCHASTIQUE DU PASSIF DES POLICES

MÉMOIRE

PRÉSENTÉ

COMME EXIGENCE PARTIELLE

DE LA MAÎTRISE EN MATHÉMATIQUES

PAR

DANAÏL DAVIDOV

NOVEMBRE 2009

UNIVERSITÉ DU QUÉBEC À MONTRÉAL

Service des bibliothèques

Avertissement

La diffusion de ce mémoire se fait dans le respect des droits de son auteur, qui a signé le formulaire *Autorisation de reproduire et de diffuser un travail de recherche de cycles supérieurs* (SDU-522 – Rév.01-2006). Cette autorisation stipule que «conformément à l'article 11 du Règlement n°8 des études de cycles supérieurs, [l'auteur] concède à l'Université du Québec à Montréal une licence non exclusive d'utilisation et de publication de la totalité ou d'une partie importante de [son] travail de recherche pour des fins pédagogiques et non commerciales. Plus précisément, [l'auteur] autorise l'Université du Québec à Montréal à reproduire, diffuser, prêter, distribuer ou vendre des copies de [son] travail de recherche à des fins non commerciales sur quelque support que ce soit, y compris l'Internet. Cette licence et cette autorisation n'entraînent pas une renonciation de [la] part [de l'auteur] à [ses] droits moraux ni à [ses] droits de propriété intellectuelle. Sauf entente contraire, [l'auteur] conserve la liberté de diffuser et de commercialiser ou non ce travail dont [il] possède un exemplaire.»

REMERCIEMENTS

Je voudrais remercier toutes les personnes qui m'ont soutenu au cours de mes études et qui ont contribué de près ou de loin à ce projet.

D'abord, un gros merci à mon directeur de maîtrise, Jean Philippe Boucher, sans qui rien de tout cela n'aurait été possible. Il m'a ouvert la voie sur le monde du travail et il a toujours su me guider et me conseiller d'une façon efficace.

Je tiens à exprimer une profonde reconnaissance à tous les professeurs de mon baccalauréat et au niveau maîtrise grâce à qui j'ai pu apprendre la statistique et l'actuariat. Je voudrais souligner les contributions particulières de Serge Alalouf et de Geneviève Lefebvre qui m'ont enseigné des outils statistiques qui me seront particulièrement utiles dans l'avenir.

Je souhaite mettre en évidence le soutien de la Fondation de l'UQAM, la Faculté des sciences de l'UQAM et Pierre J. Jeannot pour les bourses d'excellences qui m'ont permis de me consacrer pleinement à mon projet.

Finalement, j'aimerais exprimer ma gratitude à ma famille qui m'a supporté inconditionnellement tout au long de mes études.

TABLE DES MATIÈRES

TABLE DES FIGURES	ix
LISTE DES TABLEAUX	xi
RÉSUMÉ	xv
INTRODUCTION	1
CHAPITRE I	
CONCEPTS PRÉALABLES	5
1.1 Définitions	5
1.2 Exemples	7
1.3 Principes d'une réserve actuarielle	9
1.4 Disposition des données	10
1.4.1 Fichiers sinistres	10
1.4.2 Triangle de développement	11
1.4.3 Tableau en année de calendrier	13
1.5 Notations et données analysées	14
1.6 Confiance envers les données	19
1.7 Quelques concepts de probabilité	22
1.7.1 Loi binomiale	22
1.7.2 Loi normale	22
1.7.3 Loi de Poisson	23
1.7.4 Loi gamma	24
1.7.5 Modèles de régression	25
CHAPITRE II	
MÉTHODES TRADITIONNELLES	29
2.1 Chain Ladder	29
2.2 London Chain	33
2.3 Bornhuetter Ferguson	36

2.4	Sommaire des méthodes traditionnelles	39
CHAPITRE III		
	MÉTHODES STOCHASTIQUES	41
3.1	Modèle de Poisson	42
3.1.1	Bases du modèle	42
3.1.2	Principe du maximum de vraisemblance	44
3.1.3	Système balancé	48
3.1.4	Restrictions du modèle	49
3.1.5	Modélisation par unité d'exposition	49
3.1.6	Algorithme d'optimisation numérique	50
3.1.7	Autres paramétrisations	53
3.1.8	Courbe d'Hoerl	55
3.1.9	Paramètre de surdispersion	56
3.1.10	Variance des estimateurs des paramètres	60
3.1.11	Variance de la réserve	62
3.1.12	Approche bayésienne	66
3.1.13	Générer la distribution de la réserve	68
3.2	Analyse des résultats	70
3.2.1	Modèle de Poisson surdispersé	70
3.2.2	Modèle gamma	72
3.2.3	Modèle de Tweedie	73
3.2.4	Comparaison des modèles stochastiques	83
CHAPITRE IV		
	MODÉLISATION DE LA VARIANCE	87
4.1	Problématique	87
4.2	Modèles de dispersion	93
4.2.1	Estimation des rapports de variance en utilisant le principe de vraisemblance maximale	94
4.2.2	Incorporation d'une tendance dans la variance du modèle	95
4.3	Modèles linéaires généralisés doubles	97

4.4	Comparaisons entre les modèles de dispersion et les doubles GLMs	108
4.4.1	Intégration des facteurs de surdispersion dans le calcul de l'EQMP	108
4.4.2	Analyse de sensibilité du paramètre de surdispersion issu du principe de vraisemblance maximale	110
4.4.3	Analyse des résultats	112
4.5	Discussion et recherche future	112
	CONCLUSION	121
	BIBLIOGRAPHIE	123
	CHAPITRE V	
	ANNEXES	127
5.1	Méthodes traditionnelles	128
5.1.1	London Chain	128
5.1.2	Bornhuetter Ferguson	129
5.2	Méthodes Stochastiques	130
5.2.1	Modèle de Poisson surdispersé	130
5.2.2	Modèle de Poisson surdispersé - Courbe de Hoerl	133
5.2.3	Modèle gamma	134
5.2.4	Modèle Tweedie	135
5.3	Modèles de dispersion	136
5.3.1	Modèle Tweedie	136
5.4	Modèles linéaires généralisés doubles	138
5.4.1	Modèle Tweedie	138

TABLE DES FIGURES

1.1	Droite des moindres carrés passant par le nuage de points	26
1.2	Les erreurs suivent une loi normale autour de la droite	28
2.1	Droite des moindres carrés pour le nuage de points entre les périodes 2 et 3	34
2.2	Les réserves sont la somme des paiements incrémentaux prévus	38
3.1	Itérations de l'algorithme Newton-Raphson	51
3.2	Le paramètre peut varier plus lorsque la courbe est aplatie	61
3.3	Fonction de vraisemblance de p à la première itération	81
4.1	Courbe optimale de régression selon le modèle normal, Poisson ou gamma	90
4.2	Aperçu des interrelations entre les deux sous-modèles	101
4.3	Première itération du DGLM	117
4.4	2 ^e itération du DGLM	118
4.5	3 ^e itération du DGLM	119
4.6	Optimisation simultanée de la fonction de vraisemblance 4.7 pour p , alors que les autres paramètres sont estimés par le principe de la déviance . .	120

LISTE DES TABLEAUX

1.1	Pourcentage des sinistres déclarés selon la période de développement . .	9
1.2	Fichier sinistre	11
1.3	Pertes incrémentales	11
1.4	Pertes cumulatives	12
1.5	Pertes en année calendrier	13
1.6	Pertes incrémentales	15
1.7	Pertes cumulatives	15
1.8	Réserves	15
1.9	Assurance-automobile en Suisse - Encouru cumulatif	17
1.10	Assurance-automobile en Suisse - Nombre de réclamations	18
2.1	Chain Ladder Somme pondérée - Encouru Cumulatif	32
2.2	Pertes ultimes prévues : Bronhuetter Ferguson - Chain Ladder	39
2.3	Réserves selon différentes méthodes	40
3.1	Paramètres du modèle de Poisson surdispersé	71
3.2	Paramètres du modèle gamma	73
3.3	Sommaire des modèles exponentiels de la classe de Tweedie	79
3.4	Estimation de p avec le principe de vraisemblance maximale	81

3.5	Paramètres du modèle Tweedie, avec $\phi = 29348$ (déviante)	84
3.6	Sommaire des différents modèles (l'exposition est appliquée seulement au modèle de Tweedie)	85
4.1	La moyenne et la variance des coûts totaux selon différentes situations	89
4.2	La variance ($\text{Var}[S]$) selon différentes moyennes ($E[S]$), pour les modèles normal, Poisson et gamma	90
4.3	Paramètres de surdispersion selon la période, $p = 1.8111$	95
4.4	ϕ_j selon la période, $p = 1.4157$, $\phi = 564.71$, $\gamma = 0.7184$	96
4.5	Paramétrisation de V , la matrice catégorielle associée aux facteurs de surdispersion	106
4.6	Facteurs de surdispersion selon un modèle de Tweedie, $p = 1.6625$	108
4.7	Facteur de surdispersion en fonction du principe, de la colonne et du nombre d'observations ajustées, pour un modèle où $p = 1.1741$ et $w = 1$	110
4.8	Contribution au facteur de surdispersion de chaque observation	111
4.9	Réserve et EQMP du modèle de Tweedie selon différentes approches pour la surdispersion	112
4.10	Décomposition de l'EQMP selon divers modèles	113
5.1	London Chain - Encouru cumulatif	128
5.2	Bornhuetter Ferguson - Encouru cumulatif	129
5.3	Encouru incrémental observé	130
5.4	Modèle Poisson surdispersé - Encouru incrémental prévu	131
5.5	Modèle de Poisson surdispersé - Encouru cumulatif	132

5.6	Modèle de Poisson surdispersé avec courbe de Hoerl - Encouru cumulatif	133
5.7	Modèle gamma - Encouru cumulatif	134
5.8	Modèle Tweedie avec expositions, $p = 1.1741$ - Encouru cumulatif	135
5.9	Modèle Tweedie avec expositions, $p = 1.8111$ - Encouru cumulatif	136
5.10	Modèle Tweedie avec expositions, $p = 1.8111$ - paramètres	137
5.11	Modèle Tweedie avec expositions, $p = 1.6625$ - Encouru cumulatif	138
5.12	Modèle Tweedie avec expositions, $p = 1.6625$ - paramètres	139

RÉSUMÉ

Ce mémoire fait une étude détaillée des méthodes utilisées pour modéliser les réserves actuarielles en assurance de dommages. Les méthodes stochastiques utilisent des modèles linéaires généralisés qui permettent d'associer une courbe de probabilités aux pertes futures. Une analyse approfondie de la classe de modèles de Tweedie est présentée, ce qui permet d'obtenir les formules d'un large spectre de modèles. Ensuite, l'ouvrage met en évidence une différence dans la nature du risque entre la fréquence et la sévérité qui suscite la nécessité d'utiliser un modèle qui accorde plus de liberté aux facteurs de surdispersion. Deux solutions sont abordées : les modèles de dispersion, basés sur le principe du maximum de vraisemblance, et les modèles linéaires généralisés doubles, axés sur le principe de la déviance.

Mots-clés : Réserves actuarielles ; Chain Ladder ; Modèles linéaires généralisés ; Loi de Tweedie ; Déviance ; Paramètre de surdispersion ; Modèles de dispersion ; Modèles linéaires généralisés doubles.

INTRODUCTION

Cet ouvrage fait une étude détaillée des techniques les plus modernes utilisées lors de l'évaluation de la réserve d'un passif actuariel en assurance de dommages. L'objectif du travail est de capter adéquatement la nature du risque en modélisant la variance en plus de la moyenne. Lorsque l'occasion se présente, l'ouvrage propose des améliorations aux modèles déjà existants et aussi de nouvelles avenues pour le développement futur.

Une réserve est la quantité d'argent que l'assureur doit mettre de côté pour rembourser les sinistralités futures des contrats en cours. Idéalement, la réserve éponge exactement les obligations futures, sans manquer, ni en avoir de trop. En assurance non-vie, les réserves représentent, en moyenne, 75 % du total du bilan (Denuit et Charpentier, 2004). Les principes comptables exigent d'avoir une valeur inscrite au bilan qui représente les engagements en cours, mais malheureusement, cette valeur est inconnue et devra être prévue. On utilise les tendances et les données passées pour essayer de déterminer quelle valeur est la plus plausible. L'estimation des provisions pose ainsi plusieurs problèmes aux actuaires : faut-il inclure une marge de sécurité, comment tenir compte des événements atypiques (fraude, catastrophes, etc.), comment un régulateur peut-il croire qu'un montant de provision est suffisant, que se passe-t-il lorsqu'on ne provisionne pas assez, qui en sont les responsables ?

Les techniques abordées sont destinées à être utilisées dans un contexte de l'assurance de dommages. À la base, l'assurance est un engagement de l'assureur à dédommager l'assuré, lorsque diverses conditions sont réunies, en échange d'un montant d'argent fixe prédéterminé appelé *prime*. L'objectif n'est pas de prédire individuellement s'il y aura sinistre ou non, mais plutôt de prévoir collectivement l'ensemble du portefeuille. Pour

ce faire, on se base sur le principe de la loi des grands nombres qui indique que lorsqu'on répète la même expérience un grand nombre de fois, la réclamation moyenne observée par individu s'approche de son espérance mathématique.

D'un autre côté, la nature aléatoire de l'assurance fait qu'il est généralement très difficile d'accepter un seul nombre comme étant la réserve indiquée. Généralement, un intervalle de nombres pourrait être jugé acceptable. Et c'est dans cette optique qu'on propose une distribution de réserves possibles. Le risque est ainsi décrit d'une façon plus complète.

À la lumière de ce qui précède, l'actuariat est une science qui essaie de comprendre des phénomènes qui ont une nature compliquée. Puisque l'information est généralement incomplète, il arrive souvent que le jugement de l'actuaire soit indispensable à l'élaboration de la distribution de la réserve. En conséquence, on dit souvent que c'est un domaine qui est à la fois scientifique, mais aussi artistique, et c'est aussi une raison pourquoi un éventail de réponses peuvent être plausibles.

Dans le même ordre d'idées, c'est aussi un domaine où les risques sous-jacents peuvent varier beaucoup d'un contrat à l'autre. On observe une grande variabilité au niveau du paiement versé par la compagnie, mais aussi au niveau du temps avant que le sinistre soit déclaré. L'objectif est de trouver une façon de modéliser conjointement le montant et la durée avant que la réclamation soit faite. Comparativement à l'assurance-vie où la durée avant que la perte soit déclarée est négligeable, cette incertitude supplémentaire ajoute une difficulté qui rend les modèles différents et donc la théorie nécessaire est nouvelle et dissemblable avec la théorie traditionnelle de l'assurance-vie.

En bref, la nature plus aléatoire de l'assurance de dommages offre beaucoup de défis de modélisation. Les récents progrès de l'informatique, ainsi que l'accès à des

banques de données de plus en plus détaillées ont permis à cette science de faire de grands progrès. À chaque année, les découvertes se font nombreuses, et pour pouvoir garder le fil, cet ouvrage tente de réunir certains des concepts les plus récents d'une façon fluide et naturelle.

CHAPITRE I

CONCEPTS PRÉALABLES

Avant de commencer, on présente quelques concepts de base de l'assurance de dommages. On définit ce qu'est une réserve, les notations qui seront utilisées et les façons de placer les données. On présente aussi quelques exemples pratiques pour comprendre la nature différente du risque selon la ligne d'affaires. Finalement, on montre aussi quelques notions de base de la théorie des probabilités.

1.1 Définitions

Pour se familiariser avec les termes utilisés dans ce travail, on introduit quelques définitions. Tout d'abord, une *réserve*, aussi appelée *provision*, est l'argent nécessaire pour régler les engagements en cours. Ensuite, une *police* est un contrat entre l'assureur et l'assuré. De même, la *période d'exposition* est la période pendant laquelle la police est en vigueur. Un ou plusieurs sinistres peuvent survenir durant cette période, certains seront déclarés au cours de cette période, d'autres seront réclamés une ou plusieurs années après que la période soit terminée. Les méthodes adressées dans cet ouvrage supposent qu'il n'y a aucune provision pour des pertes reliées à des sinistres qui surviennent après la date d'évaluation même si le contrat est encore en vigueur. On ne considère que les sinistres qui sont déjà survenus dans le passé, qu'ils aient été déclarés ou non.

Pour distinguer ces évènements, on divise la réserve en cinq composantes majeures :

- Provision sur les sinistres encourus
- Ajustements futurs des réserves individuelles
- Sinistres survenus mais non-rapportés
- Sinistres rapportés mais non-enregistrés
- Réclamations fermées qui peuvent rouvrir ¹

La provision sur les sinistres encourus représente les obligations face à des sinistres survenus, rapportés et enregistrés. Il se peut que ces montants soient modifiés si l'estimation initiale de la perte est imparfaite. Alors, l'évolution de ces montants est considérée par le développement des sinistres encourus. Une partie importante de la réserve se caractérise par des sinistres survenus mais qui ne sont pas encore déclarés, particulièrement pour les lignes d'affaires qui ont une cadence de règlement plus longue. Une fois une perte déclarée, il y a un délai de l'ordre de quelques jours avant que cette perte soit enregistrée dans la banque de données. C'est ce qu'on appelle la réserve des sinistres rapportés mais non-enregistrés. Finalement, certains dossiers peuvent rouvrir à la suite de certains événements, et donc il se peut qu'il y ait des pertes supplémentaires qui soient encourues.

Parallèlement, il existe une réserve additionnelle qui s'occupe des frais afférents ².

1. En anglais :

Case reserve

Development on incurred claims

Incurred But Not Reported

- Claims In transit

Reopened claims

2. Loss Adjustment Expenses (LAE)

Il s'agit de dépenses de la compagnie d'assurance qui sont une conséquence des réclamations, mais qui ne sont pas des paiements en indemnités. Par exemple, les frais juridiques suite à une poursuite. On peut séparer cette réserve en deux catégories : les frais directement reliés à une réclamation particulière³ et ceux qui sont indirects⁴ (Wiser et Gardner, 1990).

1.2 Exemples

Voici un exemple fictif pour montrer concrètement l'évolution d'une réclamation. La situation se déroule dans un contexte d'assurance-automobile pour les dommages corporels.

- Le 1^{er} janvier 2004, le contrat de l'assuré entre en vigueur.
- Le 15 novembre 2004, un assuré frappe un piéton avec sa voiture.
- Le 31 décembre 2004, le contrat de l'assuré prend fin.
- Le 10 janvier 2005, le piéton commence à ressentir de violents maux de tête et de dos, conséquence de l'accident.
- Le 20 janvier 2005, une poursuite est engagée contre le conducteur par le piéton.
- Le 22 janvier 2005, l'assuré contacte son assureur pour l'avertir de la réclamation.
- Le 22 janvier 2005, le service d'indemnisation de la compagnie d'assurance évalue automatiquement la réclamation à 20 000 \$. Ce montant est classé dans la provision pour les sinistres encourus.
- Le 1^{er} mars 2005, des spécialistes de la compagnie d'assurance évaluent la réclamation à 200 000 \$. Encore une fois, ce montant est classé dans la provision pour les sinistres encourus.
- Le 30 mars 2005, le piéton refuse l'offre de 180 000 \$ de la compagnie d'assurance et va en cour.

3. Allocated Loss Adjustment Expenses (ALAE)

4. Unallocated Loss Adjustment Expenses (ULAE)

- Le 30 juin 2005, la compagnie d'assurance paie 15 000 \$ de frais légaux pour les services d'avocats.
- Le 30 mai 2006, la compagnie d'assurance paie 30 000 \$ de frais légaux pour les services d'avocats.
- Le 6 octobre 2007, la décision de la cour est une indemnité de 250 000 \$.
- Le 13 novembre 2008, le dossier est rouvert après que le piéton ait apporté un nouvel élément de preuve pour exiger des indemnités supplémentaires.
- Le 14 décembre 2008, la compagnie paie 15 000 \$ de frais légaux pour les services d'avocats.
- Le 15 décembre 2008, la cour tranche en faveur de la compagnie, aucune indemnité supplémentaire au-delà des 250 000 \$ ne sera versée.

Cet exemple montre premièrement comment les indemnités peuvent être payées après la période de couverture. Les pertes enregistrées doivent être réévaluées de 20000\$ à 200000\$ à 250000\$. Ces ajustements se font ressentir quelquefois longtemps après que le sinistre soit initialement déclaré. On note que la compagnie n'effectue l'indemnisation que le 6 octobre 2007. Par ailleurs, les frais juridiques sont classés dans la réserve des frais afférents.

Les domaines principaux de l'assurance de dommages sont distingués ici par la cadence de règlement. Lorsque le montant est plus facilement évalué, par exemple un bien courant, alors le règlement se fait sans complications, et donc la cadence de règlement est rapide. À l'inverse, quand on doit avoir recours à la justice, alors la cadence de règlement est souvent plus longue. Le tableau 1.1 résume les tendances pour divers domaines principaux (Denuit et Charpentier, 2004).

Les chiffres de la première ligne représentent l'année de développement, c'est-à-dire le temps en années avant que le sinistre soit déclaré à la compagnie d'assurance.

Domaine	1	2	3	4	5
Multi-risque habitation	55%	90%	95%	95%	96%
Automobile matériel	55%	79%	84%	89%	90%
Automobile corporel	13%	38%	50%	65%	72%
Responsabilité civile	10%	25%	35%	40%	45%

Tableau 1.1 Pourcentage des sinistres déclarés selon la période de développement

On voit que dans la première année après que le sinistre ait eu lieu, 55% des sinistres seulement sont déclarés en assurance habitation. En d'autres mots, l'assurance des biens a une cadence de règlement plus rapide que l'assurance corporelle. À l'opposé, la responsabilité civile, par exemple l'assurance de la santé de travailleurs, peut prendre beaucoup plus d'années à se résoudre car les symptômes peuvent apparaître sur le tard seulement.

1.3 Principes d'une réserve actuarielle

Peu importe le modèle ou la méthode utilisée, l'organisme qui régit les réserves actuarielles a défini des standards et principes auxquels on doit se conformer. Les principes d'estimation de la réserve selon la *Casualty Actuarial Society*, en Amérique du Nord, tels que définis dans (Actuarial Standards Board, 1991) se traduisent par :

1. Pour une date d'évaluation donnée et un groupe de risques bien défini, une provision actuarielle adéquate pour les sinistres encourus doit être basée sur des estimés dérivés à partir d'hypothèses raisonnables et de méthodes actuarielles appropriées qui jaugent les pertes à venir, qu'elles aient été déclarées ou non, pour lesquelles il existe un passif pour une date de comptabilité particulière.
2. Pour une date d'évaluation donnée et un groupe de risques bien défini, une provision actuarielle réaliste pour les frais afférents aux sinistres doit être basée sur des estimés dérivés à partir d'hypothèses raisonnables et de méthodes actuarielles

appropriées qui jaugent les pertes à venir nécessaires à l'investigation, la défense et la fermeture des dossiers, que les pertes ait été déclarées ou non, pour lesquelles il existe un passif pour frais afférents pour une date de comptabilité particulière.

3. L'incertitude inhérente à l'estimation des provisions pour les pertes à venir ainsi que pour les frais afférents implique qu'il existe un intervalle de provisions qui peuvent être acceptables. La vraie valeur des pertes totales, pour n'importe quelle date de comptabilité, ne peut être connue que lorsque tous les dossiers sont clos.
4. La réserve actuarielle la plus appropriée, parmi un intervalle de réserves plausibles, dépend de la vraisemblance relative des estimés possibles pour l'intervalle de réserves possibles, et aussi elle dépend du contexte financier dans lequel la réserve évolue et est présentée.

Ces principes indiquent qu'il faut utiliser des estimés et des modèles actuariels raisonnables pour toute réserve. De plus, puisqu'il y a une incertitude inhérente à l'estimation des pertes, il existe un intervalle de réponses possibles acceptables. Finalement, la réserve la plus appropriée dépend aussi de la confiance accordée aux estimateurs en plus du contexte financier dans lequel la réserve évolue.

1.4 Disposition des données

Tous les modèles se basent sur des tendances et des statistiques des données passées. Celles-ci peuvent être récoltées et enregistrées de diverses manières, certaines étant plus détaillées que d'autres.

1.4.1 Fichiers sinistres

Les fichiers sinistres sont très détaillés et présentent l'information de chaque réclamation individuelle, ainsi que de l'assuré qui lui correspond. Cette version est utilisée en tarification en assurance de dommages et met l'emphasis sur une analyse indi-

Police	Année	Sexe	Âge	Véhicule	...	Exposition	No. Sinistre	Montant de la réclamation
10001221	2004	M	25	1A	...	180	1	10 000
10001222	2007	M	45	2A	...	365	1	300
10001222	2005	M	38	1B	...	365	2	45 000
10001256	2004	M	19	6A	...	70	1	8 000
...

Tableau 1.2 Fichier sinistre

viduelle des polices⁵. Ces données proviennent des fichiers comptables, et les banques de données peuvent atteindre plusieurs millions de lignes, appelées aussi *observations*.

1.4.2 Triangle de développement

Comme approche alternative, on peut considérer les triangles de développement. On s'intéresse au portefeuille en général et on a donc besoin d'une partie de l'information du fichier des sinistres seulement. Un triangle classé *par année d'accident* met tous les sinistres qui surviennent dans la même année sur la même ligne, peu importe le moment où le sinistre a été déclaré. Le tableau 1.3 montre un exemple fictif.

Année	1	2	3	4
2004	100	30	10	0
2005	110	35	10	
2006	120	40		
2007	125			

Tableau 1.3 Pertes incrémentales

5. En anglais, micro-level

Année	1	2	3	4
2004	100	130	140	140
2005	110	145	155	
2006	120	160		
2007	125			

Tableau 1.4 Pertes cumulatives

Les montants du triangle 1.3 sont des pertes qui ont été déclarées à l'assureur, mais qui n'ont pas nécessairement été réglées. Les montants dans la première colonne indiquent les pertes qui sont encourues dans l'année suivant l'avènement du sinistre. Par exemple, à la fin de l'année 2004, l'assureur a eu un total de 100 \$ de pertes pour l'année 2004. En 2005, l'assureur a eu un total de 110 \$ de pertes de sinistres qui se sont passés pendant l'année 2005. Ensuite, la deuxième colonne indique des pertes supplémentaires, mais qui ont été déclarées un an en retard. Donc, dans l'exemple, il y a eu 30 \$ de pertes additionnelles de sinistres survenus pendant 2004, mais qui ont été déclarées pendant l'année 2005. Les colonnes sont appelées *période de développement* et ont la même signification que dans la section 1.2.

La même information peut être structurée en version *cumulative* telle que présentée dans le tableau 1.4. Ici, on s'intéresse au total des sommes payées jusqu'à la période indiquée. Par exemple, après 4 années, l'assureur a enregistré un total de 140 \$ de pertes survenues pendant l'année 2004. Il est à noter que pour les années anciennes, on a plus de périodes d'observation, d'où la forme (et le nom) des triangles de développement. Comparativement au fichier des sinistres, on modélise l'ensemble des assurés plutôt qu'individuellement. Il est préférable d'utiliser cette approche lorsqu'on peut aisément supposer que le groupe est homogène et que donc les différences entre les polices au niveau du risque sont négligeables. L'avantage de cette approche est qu'on peut résumer l'information sur une page seulement, et déceler à l'oeil plus facilement les tendances et les anomalies.

Année	Pertes
2004	100
2005	140
2006	165
2007	175

Tableau 1.5 Pertes en année calendrier

En supposant qu'il n'y a pas de pertes supplémentaires encourues après la quatrième période de développement, on définit les pertes ultimes comme étant les pertes cumulatives de la dernière période. De même, on peut alors dire que pour l'année 2004, $\frac{100}{140} = 71.43\%$ des pertes sont survenues lors de la première période de développement. Aussi, on obtient 21.42 % lors de la période suivante et 7.14 % lors de la troisième période.

1.4.3 Tableau en année de calendrier

Cette approche, plus condensée que les autres, permet de voir le portefeuille sous un autre point de vue. Ici, on s'intéresse au moment où les pertes sont déclarées, peu importe leur année de survenance. Pour l'exemple précédent, supposons que l'assureur s'est lancé en affaires en 2004, alors le tableau 1.5 décrit les pertes enregistrées par année de calendrier. Le lecteur s'assurera que les pertes en année de calendrier sont la somme des diagonales du triangle en année d'accident incrémental. On présente l'information de cette façon lorsqu'on veut constater l'évolution de la grandeur du portefeuille selon l'année financière. Cette approche se confond bien avec la comptabilité de caisse et elle est utilisée dans les bilans financiers.

1.5 Notations et données analysées

Toutes les méthodes sont appliquées sur les données provenant de *l'assurance-automobile en Suisse*⁶. Plus précisément, on emploie les données telles que retrouvées dans (Wüthrich et Merz, 2008). Les données, présentées en pertes cumulatives, sont à la fin de la section, et sont suivies du triangle du nombre de réclamations qui lui correspond.

L'ouvrage présent aura la même notation que dans (England et Verrall, 2002). Les pertes cumulatives observées et prévues sont notées $D_{i,j}$ et $\hat{D}_{i,j}$ respectivement, alors que les pertes incrémentales observées et prévues, $C_{i,j}$ et $\hat{C}_{i,j}$. L'indice i est réservé pour indiquer l'année d'accident (la ligne). L'indice j est réservé pour l'année de développement (la colonne). Le triangle possède I lignes et J colonnes. Dans tous les cas, les montants prévus sont en italique. Finalement, on dénote par λ_j le facteur multiplicatif pour passer de la période j à la période suivante. L'interprétation de ce facteur est expliquée dans la section 2.1.

L'objectif de toutes les méthodes qui utilisent les triangles est de remplir la partie inférieure du triangle. Ces montants représentent les pertes à venir. Les réserves sont la somme des pertes incrémentales futures. Pour un triangle où le nombre de lignes est égal au nombre de colonnes, nous avons :

$$R_i = \hat{C}_{i,J-i+1} + \dots + \hat{C}_{i,J} .$$

En supposant que les pertes ultimes sont celles de la période 4, alors on obtient les réserves du tableau 1.8. La réserve totale est la somme de la réserve de chaque année d'accident :

$$R = R_1 + R_2 + \dots + R_I .$$

6. Swiss Motor Insurance

Année	1	2	3	4
2004	$C_{1,1}$	$C_{1,2}$	$C_{1,3}$	$C_{1,4}$
2005	$C_{2,1}$	$C_{2,2}$	$C_{2,3}$	$\hat{C}_{2,4}$
2006	$C_{3,1}$	$C_{3,2}$	$\hat{C}_{3,3}$	$\hat{C}_{3,4}$
2007	$C_{4,1}$	$\hat{C}_{4,2}$	$\hat{C}_{4,3}$	$\hat{C}_{4,4}$

Tableau 1.6 Pertes incrémentales

Année	1	2	3	4
2004	$D_{1,1}$	$D_{1,2}$	$D_{1,3}$	$D_{1,4}$
2005	$D_{2,1}$	$D_{2,2}$	$D_{2,3}$	$\hat{D}_{2,4}$
2006	$D_{3,1}$	$D_{3,2}$	$\hat{D}_{3,3}$	$\hat{D}_{3,4}$
2007	$D_{4,1}$	$\hat{D}_{4,2}$	$\hat{D}_{4,3}$	$\hat{D}_{4,4}$

Tableau 1.7 Pertes cumulatives

Année	Réserve	Calcul cumulatif	Calcul incrémental
2004	R_1	$D_{1,4} - D_{1,4}$	$= 0$
2005	R_2	$\hat{D}_{2,4} - D_{2,3}$	$= \hat{C}_{2,4}$
2006	R_3	$\hat{D}_{3,4} - D_{3,2}$	$= \hat{C}_{3,3} + \hat{C}_{3,4}$
2007	R_4	$\hat{D}_{4,4} - D_{4,1}$	$= \hat{C}_{4,2} + \hat{C}_{4,3} + \hat{C}_{4,4}$

Tableau 1.8 Réserves

Une panoplie de méthodes sera abordée au cours des sections suivantes. Toutes ces techniques engendrent une réponse finale qui peut être tout aussi valide l'une que l'autre. Chaque procédé possède ses avantages et ses inconvénients, et il faut peser minutieusement le pour et le contre de chaque technique pour déterminer le petit chiffre qui sera inscrit au bilan de la compagnie.

Les tableaux 1.9 et 1.10 illustrent les données qui seront analysées tout au long de ce mémoire. Le triangle du tableau 1.9 représente les paiements effectués par une (ou plusieurs) compagnie d'assurance Suisse. On l'appelle aussi *triangle du payé*⁷. En général, le payé a un comportement plus stable que l'*encouru*⁸, défini par la somme du payé et de la provision sur les sinistres encourus. Puisque la provision sur les sinistres encourus est portée à être réévaluée tout au long de la vie d'un sinistre, elle risque de changer dans le temps, contrairement aux pertes payées par la compagnie qui, elles, ne changent pas une fois effectuées.

Pour des fins de simplification, on suppose dans ce mémoire que l'encouru est en tout point égal au payé et on emploiera sans distinction les mots *pertes*, *encouru* et *payé*. En pratique, il faut faire cette distinction et appliquer possiblement deux modèles différents lorsque les triangles de l'encouru et du payé sont jugés différents.

7. Paid Claims

8. Incurred Claims

Année	1	2	3	4	5	6	7	8	9	10	11
1	17841110	25283543	26178956	26586700	26793830	26855399	26871377	26896301	26897537	26913180	26913501
2	19519117	26175637	27117095	27272490	27341948	27379717	27433549	27544940	27587203	27613036	
3	19991172	26318655	27418832	27698481	27861135	27931135	27988013	27997894	28017550		
4	19305646	25195437	25988457	26297499	26443420	26540885	26568408	26630328			
5	18291478	24084760	24774204	25062830	25408354	25518939	25634782				
6	18832520	24573734	25155532	25404095	25510970	25605182					
7	17152710	23060996	23585802	23816258	24163162						
8	16615059	21726236	22279513	22532390							
9	16835453	21837350	22326706								

Tableau 1.9 Assurance-automobile en Suisse - Encouru cumulatif

Année	1	2	3	4	5	6	7	8	9	10	11
1	6229	3500	425	134	51	24	13	12	6	4	1
2	6395	3342	402	108	31	14	12	5	6	5	
3	6406	2940	401	98	42	18	5	3	3		
4	6148	2898	301	92	41	23	12	10			
5	5952	2699	304	94	49	22	7				
6	5924	2692	300	91	32	23					
7	5545	2754	292	77	35						
8	5520	2459	267	81							
9	5390	2224	223								

Tableau 1.10 Assurance-automobile en Suisse - Nombre de réclamations

1.6 Confiance envers les données

En assurance de dommages, on peut rencontrer une multitude de situations différentes. Parfois, la stabilité de la ligne d'affaires fait qu'il est facile de prévoir l'encours futur, alors que dans d'autres cas, c'est le désordre et on doit se fier à notre jugement pour trouver une solution adéquate. Chaque modèle possède des particularités qui le rendent plus approprié à l'une ou l'autre de ces situations. Voici les deux types de situations :

I - Les données empiriques reflètent bien la situation

La plupart des modèles offrent de meilleures performances lorsque les données sont stables et prévisibles. On peut reconnaître cette situation lorsque :

1. Le portefeuille est homogène ;
2. Le portefeuille est grand ;
3. Les conditions dans lesquelles le portefeuille évolue sont stables ;
4. Il n'y a pas d'événements extrêmes ;
5. Les variances du nombre de réclamations et de la sévérité sont petites.

Prenons l'exemple de l'assurance automobile pour les dommages matériels. Pour qu'un triangle soit jugé homogène, il faut qu'il soit composé de risques similaires. En d'autres mots, les conducteurs doivent avoir une fréquence de sinistres semblables, et utilisent des voitures de valeur comparable. Un portefeuille composé de milliers de polices possède un volume suffisant pour pouvoir utiliser la théorie des grands nombres sans aucune réticence. Pour pouvoir affirmer que les conditions sont stables, il faut que les lois concernant le traitement des sinistres restent inchangées au cours des années observées. Aussi, l'entreprise doit conserver une stabilité dans sa politique des priorités quant aux traitements des dossiers selon l'importance du sinistre. De plus, il ne doit pas y avoir de catastrophes naturelles telles qu'un tremblement de terre important, ou un ouragan immense, sinon il y aurait un nombre de réclamations bien au-delà du nombre

habituel.

L'assurance des biens est un domaine où la sévérité est moins variable que celle de l'assurance invalidité. En effet, on peut évaluer rapidement la valeur du bien (comparativement à d'autres domaines) et si les individus ont une valeur assurée semblable, la sévérité qui leur est associée le sera aussi. À l'inverse, pour l'assurance de personne, la blessure peut être légère, ou grave, et les compensations pour remplacer le revenu de la personne dépendent beaucoup de l'individu. Donc l'assurance-habitation et l'assurance automobile au niveau matériel sont des domaines propices pour faire confiance aux données.

On peut penser à sectionner les éléments d'un portefeuille en plusieurs portefeuilles plus petits et indépendants. Ceux-ci seront vraisemblablement plus homogènes, mais seront alors plus petits. Ainsi, en essayant de combler un manque, on en crée un autre (Berquist et Sherman, 1977).

Peu importe le modèle qu'on applique, une partie de la modélisation est issue de notre jugement. Même lorsqu'on considère un modèle entièrement objectif, il faut se rappeler que la construction du modèle possède une partie subjective. En effet, on a le choix d'utiliser un modèle particulier (par rapport à d'autres) ou bien d'en comparer plusieurs et de choisir plusieurs solutions finales.

En pratique, il arrive souvent que certaines conditions ne soient pas entièrement satisfaites. On assiste alors à un deuxième type de situation.

II - On ne peut pas faire confiance aux données empiriques

On reconnaît ces situations s'il y a au moins un de ces éléments :

1. Le portefeuille est hétérogène ;
2. Le portefeuille est trop petit ;
3. Les conditions dans lequel le portefeuille évolue sont instables ;
4. Il y a des événements extrêmes ;
5. Les variances du nombre de réclamations et de la sévérité sont grandes.

Lorsque la situation l'exige, il arrive que l'actuaire décide que certains nombres doivent être ajustés à la hausse ou à la baisse. Par exemple, dans un contexte déterministe (chapitre 2), la méthode de Bornhuetter-Ferguson permet de fixer les pertes ultimes à des valeurs prédéterminées. Dans un contexte stochastique (chapitre 3), on peut se référer aux modèles bayésiens qui offrent aussi une flexibilité similaire (section 3.1.12).

En pratique, ces procédés sont utilisés lorsqu'on part une nouvelle ligne d'affaires, ou bien dans les domaines où on assure le corporel et la responsabilité civile. La sévérité est connue au terme d'une poursuite judiciaire et possèdent une grande variance. Lorsque qu'il y a un changement de loi dans l'industrie, il faut aussi adapter le modèle, les données passées pouvant mal refléter les tendances présentes et futures. Finalement, lorsque le portefeuille est affecté par une catastrophe de grande envergure, il est naturel d'ajuster les tendances en conséquence (Clark, 1996). Par ailleurs, on peut avoir recours à la réassurance pour essayer de limiter l'étendue des pertes, et ainsi diminuer la variance de la sévérité et du nombre de réclamations.

Il arrive aussi d'observer des paiements négatifs lors de périodes matures. En effet, ceux-ci peuvent survenir lorsqu'on revoit à la baisse des paiements à venir ou que certains cas se règlent en cour au profit de l'assureur. Plusieurs modèles sont très mal adaptés à cette situation, et c'est pourquoi il arrive souvent qu'on ait à ajuster manuellement certains paramètres du modèle.

1.7 Quelques concepts de probabilité

Les modèles stochastiques utilisent diverses distributions qui seront présentées ici. Ensuite, quelques explications concernant les modèles linéaires généralisés seront ajoutées.

1.7.1 Loi binomiale

La loi binomiale est l'une des premières loi de probabilités qui a été développée. Concrètement, elle consiste à compter le nombre de succès d'une expérience qui compte n essais indépendants qui ont chacun une probabilité p d'être considéré comme un succès. La fonction de probabilité, ainsi que son espérance et sa variance sont :

$$\Pr(Y = k \mid n; p) = \binom{n}{k} p^k (1-p)^{n-k} \quad , \quad k \in \{0, 1, \dots, n\} \quad , \quad 0 < p < 1 \quad , \quad n \in \mathbb{N}^* \quad ,$$

$$\mathbb{E}[Y] = np \quad , \quad \text{Var}[Y] = np(1-p) \quad .$$

1.7.2 Loi normale

La loi normale est une distribution très utilisée en statistique. Elle a été développée par le mathématicien Abraham De Moivre⁹ en 1733 afin d'approximer une loi binomiale qui possède un paramètre n très grand. Cette loi a ensuite été popularisée par Gauss¹⁰. On l'appelle aussi loi gaussienne, ou *courbe en cloche*¹¹. La fonction de densité, ainsi que son espérance mathématique et sa variance sont :

$$f(y \mid \mu; \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{y-\mu}{\sigma}\right)^2} \quad , \quad y \in \mathbb{R} \quad , \quad \mu \in \mathbb{R} \quad , \quad \sigma > 0 \quad ,$$

$$\mathbb{E}[Y] = \mu \quad , \quad \text{Var}[Y] = \sigma^2 \quad .$$

9. Abraham De Moivre (1667-1754) mathématicien français né à Vitry-le-François

10. Carl Friedrich Gauss (1777-1855) physicien, mathématicien et astronome allemand

11. Bell curve

Deux propriétés importantes de cette loi seront considérées au cours de cet ouvrage. Premièrement, le théorème limite central indique que la somme d'un grand nombre de variables aléatoires indépendantes de même moyenne et de variance finie peut être approximée par une loi normale. Ainsi, il suffit de connaître l'espérance et la variance des variables aléatoires et que le nombre d'observations soit suffisamment grand, mais fini, pour pouvoir appliquer cette approximation.

La seconde propriété est que le paramètre de la moyenne de la loi normale n'apparaît pas dans la formule de la variance de la loi normale. Ainsi, pour les régressions, lorsque la moyenne change, la variance reste la même. On dit alors que la variance est constante, ou que la régression a la propriété d'homoscédasticité.

1.7.3 Loi de Poisson

La loi de Poisson a été publiée pour la première fois par Siméon-Denis Poisson en 1838. C'est une loi discrète (qui ne prend que des valeurs entières positives) qui avait pour objectif de modéliser une variable qui compte un nombre d'occurrences d'un événement dans le temps. Par exemple, cette loi a été utilisée pour modéliser l'arrivée de bateaux dans un port ou bien le nombre d'accidents dus aux coups de pieds de chevaux dans les armées prussiennes dans le 19^e siècle. La fonction de probabilité, ainsi que l'espérance et la variance d'une loi de Poisson sont :

$$\begin{aligned}\Pr(Y = k \mid \lambda) &= \frac{e^{-\lambda} \lambda^k}{k!}, \quad k \in \mathbb{N}, \lambda > 0 \\ E[Y] &= \lambda, \quad \text{Var}[Y] = \lambda.\end{aligned}$$

La loi de Poisson est la limite d'une loi binomiale où le nombre d'essais (n) tend vers l'infini, alors que la probabilité de succès (p) tend vers 0 et que le produit des deux (np) converge vers un réel (λ), qui est l'espérance de la loi de Poisson. De plus, son caractère discret et positif lui permet d'être mieux adaptée pour modéliser une variable de comptage.

La loi de Poisson possède une variance qui dépend de sa moyenne. En fait, elle est égale à sa moyenne. Donc dans les régressions, lorsque la moyenne augmente, la variance augmente de façon proportionnelle. Lorsque la variance n'est pas constante d'une observation à une autre, on dit que c'est une propriété d'hétéroscédasticité.

1.7.4 Loi gamma

La loi gamma est une loi continue qui est souvent employée pour modéliser la sévérité des pertes. La fonction de densité, l'espérance et la variance sont :

$$f(y \mid \alpha; \tau) = \frac{\tau^\alpha}{\Gamma(\alpha)} y^{\alpha-1} e^{-\tau y} \quad , \quad y > 0$$

$$E[Y] = \frac{\alpha}{\tau} \quad , \quad \text{Var}[Y] = \frac{\alpha}{\tau^2} \quad ,$$

avec $\Gamma(\alpha) = \int_0^\infty y^{\alpha-1} e^{-y} dy$. On peut reparamétriser cette loi de la façon suivante pour identifier clairement les paramètres qui agissent sur la moyenne et la variance :

$$\alpha = \frac{1}{\phi} \quad , \quad \tau = \frac{1}{\mu\phi} \quad ,$$

$$E[Y] = \frac{\alpha}{\tau} = \frac{1/\phi}{1/(\mu\phi)} = \mu \quad , \quad \text{Var}[Y] = \frac{\alpha}{\tau^2} = \frac{1/\phi}{1/(\mu\phi)^2} = \phi\mu^2 \quad .$$

Ainsi, on remarque que la variance de la loi gamma dépend du carré de sa moyenne ; c'est une caractéristique qui sera importante dans un contexte de régression.

Par ailleurs, la loi gamma est une généralisation de la loi du Chi-deux, qui elle-même est reliée à la loi normale. En effet, soit X_i suit une loi normale centrée réduite. Alors, $Z_i = X_i^2$ qui suit une loi Chi-deux avec un degré de liberté. Ensuite, soit $Z = \sum_{i=1}^n Z_i$, la somme de n variables aléatoires indépendantes qui suivent chacune une loi Chi-deux avec un degré de liberté, alors Z suit une loi Chi-deux avec n degrés de liberté. Finalement, il se trouve que $Z \sim \Gamma(\frac{n}{2}; \frac{1}{2})$: la loi du Chi-deux est un cas

particulier de la loi gamma.

1.7.5 Modèles de régression

Les modèles stochastiques discutés dans ce mémoire se basent sur la théorie des régressions. C'est un outil statistique très puissant qui permet de déceler les tendances et les relations entre les variables. Supposons que nous avons un vecteur de variables X qui correspond à des observations en des lieux précis au début de la période, et Y le vecteur d'observations des mêmes endroits à la fin de la période. L'objectif est de déterminer si les valeurs de Y dépendent d'une quelconque façon des valeurs de X . D'abord, on peut placer ces données dans un nuage de points pour voir si on ne peut déceler une tendance à l'oeil. Ensuite, on peut faire passer une droite qui a pour but de passer le mieux possible dans ce nuage de points. On dit alors que X est la variable explicative et Y la variable expliquée. La forme de la courbe est la nature de la relation entre les deux variables.

À l'origine, une des premières méthodes pour trouver une relation entre les deux variables est la droite des moindres carrés. Selon les historiens, cette méthode a été trouvée indépendamment par Carl Ferdrich Gauss et Adrien-Marie Legendre¹². L'objectif est de faire passer la meilleure droite dans le nuage de points. Bien entendu, elle ne pourra pas passer par tous les points, mais elle essaiera d'être le plus près possible en minimisant une fonction de pénalité qui, dans le cas de la figure 1.1, est la somme des carrés de l'erreur.

Pour une valeur X donnée, la droite définie par $\hat{Y} = aX + b$ tente d'approcher la valeur expliquée, Y . La différence entre Y et \hat{Y} est ce qu'on appelle l'erreur. On

12. Adrien-Marie Legendre (1752-1853) mathématicien français né à Paris

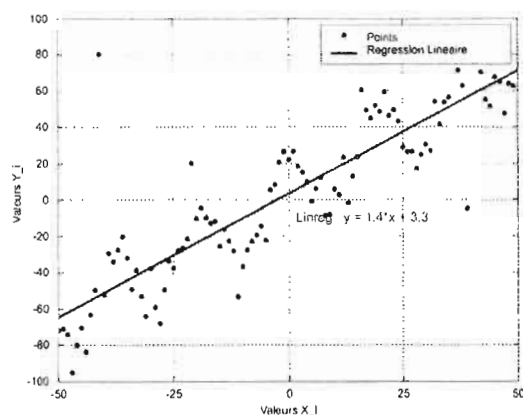


Figure 1.1 Droite des moindres carrés passant par le nuage de points

minimise alors :

$$S^2 = \sum_{i=1}^n e_i^2 ,$$

où n est le nombre d'observations et e_i , l'erreur de la i^e observation.

On peut dire qu'on régresse Y sur X . Le terme *régression* peut paraître curieux. Il a été introduit par Sir Francis Galton¹³ en 1855. Dans une première étude, il avait remarqué que les plantes provenant de plantes ayant beaucoup produit de graines ont tendance à produire moins de graines plus tard. D'autre part, les plantes ayant produit moins de graines ont tendance à en produire plus les périodes suivantes. Ainsi, dans les deux cas, il a noté un retour vers la moyenne. Galton y voyait une régression, en ce sens qu'on observait la disparition d'un caractère distinctif.

Les techniques ont beaucoup évolué depuis. Tout d'abord, il existe ce qu'on appelle les régressions multiples. L'idée est d'utiliser de l'information provenant de plusieurs

13. Sir Francis Galton (1822-1911) anthropologue britannique, beau-frère de Charles Darwin

sources (vecteurs explicatifs) pour essayer d'expliquer la variable dépendante, notée Y . Par exemple, on peut chercher à modéliser la grandeur d'une plante en fonction de la quantité de pluie, la qualité de la terre, la quantité de soleil et la quantité d'engrais. On modélise conjointement toute cette information. On obtient alors un nuage de points en cinq dimensions, quatre pour les variables explicatives, notées X_i , et une dimension pour la variable expliquée. De nouveau, on essaie de faire passer l'hyperplan qui passe le plus près du nuage de points. La formule de l'hyperplan est alors :

$$\hat{Y}_i = a + b_1 X_{i1} + b_2 X_{i2} + b_3 X_{i3} + b_4 X_{i4} .$$

Comme chaque plante possède ses propres caractéristiques au départ, la valeur prévue lui sera particulière. En remarquant que la série d'additions est un produit vectoriel, on peut écrire le vecteur des valeurs prévues sous forme matricielle :

$$\hat{Y} = X\beta ,$$

où $\beta = (a \ b_1 \ b_2 \ b_3 \ b_4)'$ et X la matrice catégorielle qui contient les informations explicatives, et \hat{Y} sont les valeurs prévues. La statistique à minimiser reste la même.

En utilisant le principe de vraisemblance maximale, on découvre que le principe des moindres carrés est strictement équivalent à supposer que les observations sont des réalisations d'une loi normale centrée autour des vraies moyennes prévues ($E[Y|X]$), mais qui sont estimées par \hat{Y} . Par ailleurs, la différence entre la *vraie* moyenne et la moyenne estimée est ce qu'on appelle l'erreur d'estimation ; ce concept est approfondi dans la section 3.1.11.

D'autre part, on peut penser à utiliser une autre loi pour expliquer la variable réponse. Par exemple, pour des données de comptage, il est naturel de penser à la loi de Poisson, qui possède un domaine discret. Si on change la loi, alors il faudra minimiser une autre statistique qui lui correspond. Ces concepts sont détaillés dans la section 3.1.

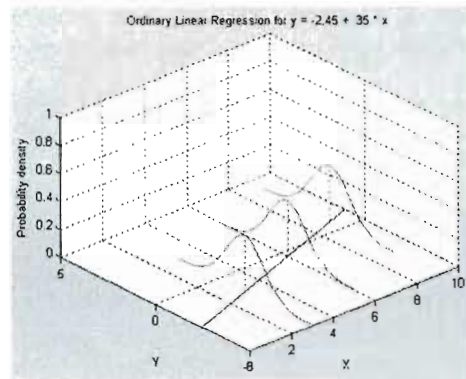


Figure 1.2 Les erreurs suivent une loi normale autour de la droite

On peut aussi faire passer une courbe au lieu d'une droite dans le nuage de points. Il suffit alors de poser la transformation suivante :

$$\hat{Y} = g(X\beta)$$

où g est une fonction monotone continue et dérivable, appelée aussi *fonction de lien*¹⁴. Lorsqu'on généralise la régression linéaire classique en utilisant une ou plusieurs caractéristiques mentionnées ci-haut, alors nous avons un modèle appelé *Modèle linéaire généralisé*¹⁵ (Nelder et Wedderburn, 1989).

14. Link function

15. Generalized Linear Model (GLM)

CHAPITRE II

MÉTHODES TRADITIONNELLES

Cette section se caractérise par des méthodes qui sont basées sur des algorithmes intuitifs qui sont les premiers à avoir été employés par les actuaires. On essaie de trouver une tendance élémentaire sans utiliser des concepts statistiques avancés. Les approches traditionnelles sont généralement plus simples et faciles à calculer. Elles peuvent être présentées sans peine à des gens qui n'ont pas de connaissances statistiques. Ainsi, les réserves peuvent être calculées rapidement, ce qui nous permet d'avoir un bon point de départ auquel seront comparées les réponses plus élaborées des méthodes stochastiques. De nos jours, la grande majorité des techniques utilisées en pratique sont les méthodes traditionnelles. Les livres de (Denuit et Charpentier, 2004) et (Wüthrich et Merz, 2008) sont des sources très détaillées sur la modélisation des réserves actuarielles.

2.1 Chain Ladder

Cette approche est simple et intuitive. On ne considère que le triangle des coûts encourus. L'idée est d'observer comment l'encours cumulatif évolue d'une période de développement à l'autre. Par exemple, s'il n'y a pas eu de coûts incrémentaux supplémentaires dans le passé pour passer d'une période à l'autre, alors, il n'y a pas de raisons que cette tendance ne soit pas répétée dans l'avenir. De même, si on observe que pour passer d'une période à l'autre, l'encours cumulatif augmente d'un certain pourcentage, alors le même phénomène devrait se reproduire pour les années d'accident futures. Les pourcentages

d'augmentation sont appelés *facteurs multiplicatifs*, ou bien *facteurs de déroulement*, ou encore *facteurs de développement*¹. Ils sont notés λ_k où k représente le passage de la période k à la période $k + 1$.

Dans chaque cas, afin de maximiser la précision de la méthode, on utilise le plus d'information possible. Il est clair que les périodes de développement jeunes ont plus d'information disponible. Par exemple, nous avons 9 années d'accident disponibles pour calculer le premier facteur de développement. À l'inverse, pour la 10^e période, nous n'avons que la première année d'accident qui possède l'information de la transition avant et après la période.

Somme pondérée

Le facteur de déroulement issu de la somme pondérée se calcule en divisant la somme de l'encouru cumulatif de la période suivante par la somme de l'encouru cumulatif de la période précédente, pour toutes les années d'accident disponibles.

Pour trouver le j^e facteur de développement, on utilise l'équation suivante :

$$\lambda_j = \frac{\sum_{i=1}^{n-j} D_{i,j+1}}{\sum_{i=1}^{n-j} D_{i,j}}, \quad j = 1, \dots, n-1,$$

où n représente le nombre d'années d'accident du triangle. Pour les triangles tronqués comme dans l'exemple du tableau 2.1, les indices des sommes doivent être légèrement ajustés pour être sûr de prendre toutes les données à notre disposition, mais le principe reste le même.

Une fois les facteurs de déroulement calculés, on trouve l'encouru prévu en multipliant celui de la période précédente par le facteur multiplicatif qui lui correspond,

1. Loss Development Factors (LDF)

comme proposé dans l'équation suivante :

$$\hat{D}_{i,j+1} = \lambda_j D_{i,j}$$

On obtient ainsi les résultats du tableau 2.1. Dans notre illustration empirique, le triangle est incomplet à une extrémité, car il manque les données $D_{10,1}$, $D_{11,1}$ et $D_{10,2}$. On dit aussi qu'il est *tronqué*. La formule des facteurs de déroulement doit alors être ajustée pour tenter de récolter le maximum d'information du triangle :

$$\lambda_j = \frac{\sum_{i=1}^{\min(n-j, n^*)} D_{i,j+1}}{\sum_{i=1}^{\min(n-j, n^*)} D_{i,j}}, \quad j = 1, \dots, n-1,$$

où n est le nombre d'années d'accident si le triangle n'est pas tronqué, et n^* est le nombre d'années d'accident dont l'information est disponible. Dans l'exemple du tableau 2.1, nous avons $n = 11$ et $n^* = 9$.

Année	1	2	3	4	5	6	7	8	9	10	11
1	17841110	25283543	26178956	26586700	26793830	26855399	26871377	26896301	26897537	26913180	26913501
2	19519117	26175637	27117095	27272490	27341948	27379717	27433549	27544940	27587203	27613036	27613365
3	19991172	26318655	27418832	27698481	27861135	27931135	27988013	27997894	28017550	28038878	28039213
4	19305646	25195437	25988457	26297499	26443420	26540885	26568408	26630328	26650729	26671017	26671335
5	18291478	24084760	24774204	25062830	25408354	25518939	25634782	25683789	25703465	25723032	25723339
6	18832520	24573734	25155532	25404095	25510970	25605182	25656698	25705747	25725440	25745023	25745330
7	17152710	23060996	23585802	23816258	24163162	24234669	24283428	24329852	24348490	24367025	24367316
8	16615059	21726236	22279513	22532390	22703663	22770851	22816664	22860284	22877797	22895212	22895485
9	16835453	21837350	22326706	22566221	22737751	22805040	22850922	22894608	22912147	22929588	22929862

Période j	1	2	3	4	5	6	7	8	9	10
λ_j	1,328	1,03	1,011	1,008	1,003	1,002	1,002	1,001	1,001	1

Tableau 2.1 Chain Ladder Somme pondérée - Encouru Cumulatif

Commentaires et critiques de la méthode

La méthode Chain Ladder est certainement la plus utilisée dans l'industrie. Elle obtient de bons résultats surtout lorsque les hypothèses de la section 1.6 sont respectées. Sa simplicité lui octroie un caractère objectif, ce qui permet d'obtenir un consensus plus facilement entre les différents actuaires qui l'utilisent. En général, ces deux qualités suffisent à inciter à faire une approximation préliminaire à l'aide de la méthode Chain Ladder pour avoir une idée générale du montant de la réserve.

À l'inverse, si les hypothèses ne sont pas soutenues par les données, les prévisions de la méthode risquent d'être erronées. Une grande perte exceptionnelle fera augmenter le taux de façon démesurée. Par exemple, si le total des paiements pour la première période est exceptionnellement deux fois plus élevé que d'habitude, les pertes ultimes prévues doubleront. Cette situation peut arriver dans les lignes d'affaires à développement lent, dans lesquelles le premier paiement est plutôt marginal par rapport à l'ensemble des paiements à venir.

D'autre part, il s'avère que pour les périodes avancées, les modèles à ailes relevées captent parfois mieux la nature du risque. Ces modèles attribuent une importance et une probabilité plus grandes aux événements inhabituels ou extrêmes. C'est dans cette intention qu'en pratique, il arrive d'ajouter une composante à queue épaisse à partir d'une certaine période de développement (Mack, 1999).

2.2 London Chain

Cette méthode est plus élaborée et plus flexible que la méthode Chain Ladder. En plus de considérer une tendance multiplicative entre les périodes (les facteurs λ_j), on suppose qu'il y a aussi une tendance incrémentale (les facteurs α_j). La méthode London Chain est une généralisation de la méthode Chain Ladder car si la tendance

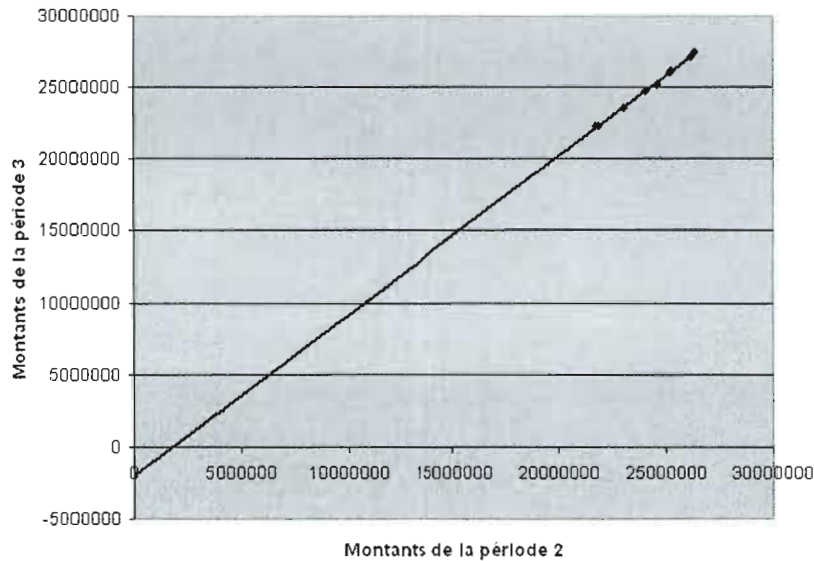


Figure 2.1 Droite des moindres carrés pour le nuage de points entre les périodes 2 et 3

incrémentale est nulle, on peut obtenir la méthode Chain Ladder, dépendamment de la façon d'estimer la droite optimale.

Pour déceler les tendances multiplicatives et additives, on utilise les modèles de régression. Pour la transition j entre les périodes j et $j + 1$, on utilise les pertes cumulatives des périodes correspondantes, comme pour la méthode Chain Ladder. Ces deux vecteurs forment un nuage de point, dans lequel on introduit la meilleure droite. La pente de la droite représente l'effet multiplicatif, alors que l'ordonnée à l'origine, l'effet incrémental. La figure 2.1 représente le nuage de points et la droite entre les périodes 2 et 3. Pour déterminer les paramètres de la droite, on minimise par convention la statistique de la somme des carrés de l'erreur (section 1.7.5).

Par exemple, pour passer de la période 2 à la période 3, on regroupe les données de la deuxième colonne (25283543, ..., 21837350) dans le vecteur D_j , et les données de

la troisième colonne (26178956 , ... , 22326706) dans le vecteur D_{j+1} . Ensuite, on fait une régression des moindres carrés de la droite qui passe par les points (D_j, D_{j+1}) . La formule de la droite des moindres carrés est la suivante :

$$\lambda_j = \frac{\frac{1}{n-j} \sum_{i=1}^{n-j} D_{i,j} D_{i,j+1} - \bar{D}_{i,j}^{(j)} \bar{D}_{i,j+1}^{(j)}}{\frac{1}{n-j} \sum_{i=1}^{n-j} D_{i,j}^2 - \left(\bar{D}_{i,j}^{(j)}\right)^2}, \alpha_j = \bar{D}_{i,j+1}^{(j)} - \lambda_j \bar{D}_{i,j}^{(j)},$$

$$\bar{D}_{i,j}^{(s)} = \frac{1}{n-j} \sum_{i=1}^{n-s} D_{i,j} \text{ et } \bar{D}_{i,j+1}^{(s)} = \frac{1}{n-s} \sum_{i=1}^{n-j} D_{i,j+1},$$

où λ_j représente l'effet multiplicatif et α_j , l'effet incrémental entre les périodes j et $j + 1$. Ainsi, le modèle s'exprime comme :

$$\hat{D}_{i,j+1} = \lambda_j D_{i,j} + \alpha_j.$$

L'encours prévu et les facteurs de développement sont présentés dans le tableau 5.1 (p.128) en annexe. Notons que α_{10} ne peut être calculé car on cherche à la fois une tendance multiplicative et additive alors qu'il n'y a qu'un seul couple d'observations. Alors, la convention veut que l'on ne considère que l'effet multiplicatif pour cette dernière période de développement.

Somme toute, il est préférable d'utiliser le London Chain seulement s'il y a de bonnes raisons de croire qu'il y a un facteur incrémental en plus d'un facteur multiplicatif, selon la nature de la ligne d'affaire et des connaissances particulières que possède l'actuaire. Pour que le modèle soit accepté, il faut que l'ajustement incrémental soit suffisamment grand. En regardant le graphique 2.1, on note que la droite ne passe pas exactement par l'origine, et donc il faut juger si l'ordonnée à l'origine semble assez éloignée de la valeur 0.

Finalement, ce modèle, tout comme le modèle Chain-Ladder, devrait être considéré seulement lorsque la ligne d'affaires est stable et prévisible.

2.3 Bornhuetter Ferguson

L'objectif primaire de la méthode est d'assurer une meilleure stabilité de l'estimation des réserves. Ceci est particulièrement vrai pour les réserves des années d'accident récentes où les réserves dépendent trop des premiers paiements lorsque la méthode Chain Ladder (section 2.1) est utilisée. La méthode proposée ici se décompose en trois étapes :

1. Déterminer les pertes ultimes espérées ;
2. Calculer les taux de déroulement en utilisant la méthode Chain-Ladder ;
3. Déterminer les proportions du montant ultime qu'il reste à payer.

Les pertes ultimes espérées peuvent être calculées selon le nombre de polices vendues, les primes reçues, ou bien établies directement à partir du jugement de l'actuaire. Dans tous les cas, l'information peut provenir d'une source extérieure aux données du triangle.

La méthode Bornhuetter Ferguson utilise une procédure analogue à celle de la méthode Chain Ladder pour le calcul des taux de déroulement, pour lesquels on suppose un lien multiplicatif. Ensuite, ceux-ci sont utilisés avec les pertes ultimes pour extrapoler dans le futur. Autrement dit, la structure des facteurs de déroulement est utilisée pour exprimer l'encours cumulé comme un pourcentage des pertes ultimes. Par exemple, pour la méthode Chain Ladder, nous avons :

$$D_{i,J} = D_{i,j} \cdot \prod_{k=j}^J \lambda_k , \quad (2.1)$$

où $\prod_{k=j}^J \lambda_k$ représente le produit des facteurs de déroulement jusqu'à la période ultime et que $D_{i,J}$ est l'encours ultime alors que $D_{i,j}$ est l'encours à la période de déroulement j . Ensuite, rappelons que la réserve de l'année d'accident i à partir de la période de développement j , notée $R_i^{(j)}$, est, par définition, la différence entre l'encours ultime et l'encours à la période de déroulement j :

$$R_i^{(j)} = D_{i,J} - D_{i,j} , \quad (2.2)$$

comme illustré dans la figure 2.2. En combinant les équations 2.1 et 2.2, on obtient :

$$R_i^{(j)} = D_{i,j} \cdot \left(1 - \frac{1}{\prod_{k=j}^J \lambda_k} \right) . \quad (2.3)$$

Pour alléger la notation de cette section, utilisons $\beta_j = \frac{1}{\prod_{k=j}^J \lambda_k}$. L'équation 2.3 devient alors :

$$R_i^{(j)} = D_{i,j} \cdot (1 - \beta_j) . \quad (2.4)$$

Le terme $(1 - \beta_j)$ représente le pourcentage des pertes qui seront encourues dans l'avenir. La méthode Bornhuetter-Ferguson suppose que ce pourcentage sera appliqué à l'estimé des pertes ultimes déterminées *a priori*, noté $D_{i,j}$. À l'inverse, la méthode Chain Ladder suppose que ce pourcentage sera utilisé pour les pertes ultimes (encore $D_{i,j}$) qui ont été estimées à partir des données disponibles à ce jour, tel qu'illustré dans l'équation 2.1. En d'autres mots, ces deux méthodes fonctionnent d'une façon similaire à un bolide qui doit parcourir une piste. Dans les deux cas, le pourcentage de la route à franchir est le même, la différence est que l'un provient d'une piste d'une longueur prédéterminée, alors que pour l'autre, la longueur de la piste est calculée à partir des données. Finalement, notons que les pertes ultimes cumulatives prévues $\hat{D}_{i,j}$ sont calculées en additionnant les réserves prévues par l'équation 2.4 aux pertes cumulatives observées à ce jour comme illustré dans l'équation suivante :

$$\hat{D}_{i,j} = D_{i,j} + D_{i,j} \cdot (1 - \beta_j) . \quad (2.5)$$

En utilisant les équations (2.1), (2.2) et (2.3), on en déduit que $\hat{D}_{i,j} = D_{i,j}$ seulement lorsque l'estimé des pertes ultimes déterminées *a priori* est exactement celui qui est proposé par la méthode Chain Ladder. Quant à la méthode Bornhuetter-Ferguson, on suppose que peu importe le passé, une fois les facteurs de déroulement calculés, les pertes incrémentales prévues seront toujours une proportion de l'estimé des pertes ultimes déterminées *a priori*, ce qui explique cette différence.

Pour trouver les pertes incrémentales prévues pour une année d'accident et une période donnée, on utilise la même approche. D'abord, on trouve les réserves à partir du

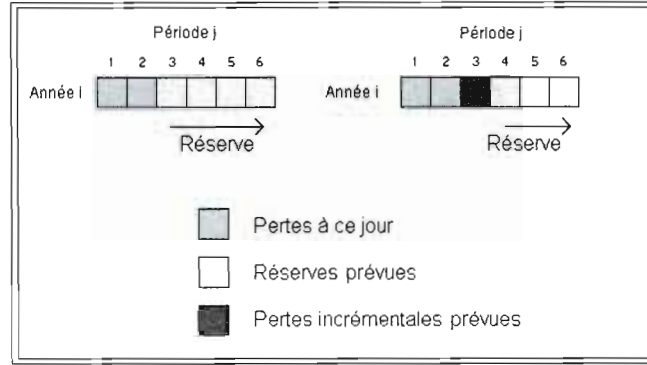


Figure 2.2 Les réserves sont la somme des paiements incrémentaux prévus

début de la période jusqu'à la période ultime. Ensuite, on trouve les réserves à partir de la fin de la période ciblée jusqu'à la période ultime. La différence entre les deux réserves représente les pertes incrémentales prévues, notées $C_{i,j}$, comme illustré dans la figure 2.2. Ainsi,

$$\begin{aligned}
 C_{i,j} &= R_i^{(j)} - R_i^{(j+1)} \\
 &= (D_{i,J}(1 - \beta_j)) - (D_{i,J}(1 - \beta_{j+1})) \\
 &= D_{i,J}(\beta_{j+1} - \beta_j)
 \end{aligned}$$

Pour des fins purement démonstratives, supposons que les estimés *a priori* des pertes ultimes ont été choisies arbitrairement et sont illustrés dans le tableau 2.2. Les réserves calculées sont disponibles dans le tableau 5.2 (p.129) en annexes.

Cette approche fait partie au même titre que la méthode Chain-Ladder un outil de base qui peut servir à faire un estimé rapide et facile de la réserve avant d'aborder des méthodes plus complexes. En bref, cette méthode est privilégiée dans les situations où le modèle Chain-Ladder capte mal la nature du risque car elle permet d'incorporer facilement de l'information extérieure issue du jugement de l'actuaire.

Année	Pertes ultimes prévues (BF)	Pertes ultimes prévues (CL)
1	27000000	26913501
2	27500000	27613365
3	28000000	28039213
4	27000000	26671335
5	26000000	25723339
6	25500000	25745330
7	24000000	24367316
8	23000000	22895485
9	23000000	22929862

Tableau 2.2 Pertes ultimes prévues : Bronhuetter Ferguson - Chain Ladder

2.4 Sommaire des méthodes traditionnelles

Généralement, plusieurs méthodes sont appliquées pour analyser un triangle de données. La variété des prédictions selon les divers modèles nous donne un portrait plus complet de ce que peut être la vraie réserve. Il est facile de comparer les méthodes complètement objectives telles que la Chain Ladder. Il est plus difficile de comparer les méthodes avec connaissances *a priori* car le résultat final dépend largement des hypothèses initiales. Le tableau 2.3 indique les réserves pour toutes les années d'accident selon la méthode utilisée.

Il est immédiatement apparent que la méthode London Chain propose des réserves élevées, particulièrement pour les années d'accidents récentes. Cette différence est due à des facteurs de déroulement incrémentaux sensiblement loin de 0 pour les périodes 2 à 4. Puisque le modèle Chain Ladder n'admet pas de lien incrémental, les deux techniques diffèrent en ce point. En ce qui concerne la méthode de Bornhuetter Ferguson, on constate que les réserves sont très semblables puisque les estimés ultimes espérés sont

Année	Chain Ladder	London Chain	Bornhuetter Ferguson
1	- \$	- \$	- \$
2	329 \$	329 \$	328 \$
3	21 663 \$	32 526 \$	21 632 \$
4	41 007 \$	16 195 \$	41 512 \$
5	88 557 \$	58 296 \$	89 509 \$
6	140 148 \$	140 012 \$	138 813 \$
7	204 154 \$	271 835 \$	201 076 \$
8	363 095 \$	702 692 \$	364 753 \$
9	603 156 \$	959 604 \$	605 001 \$
Total	1 462 108 \$	2 181 489 \$	1 462 624 \$

Tableau 2.3 Réserves selon différentes méthodes

très proches de ceux de la méthode Chain Ladder.

Somme toute, chaque méthode traditionnelle propose une réserve prévue particulière. Plusieurs réponses peuvent être acceptables. Malheureusement, aucune de ces méthodes n'indique le degré de précision des réserves. Quel montant faut-il mettre de côté pour être sûr à 95 % de pouvoir régler la totalité des contrats ? Pour répondre à ce type de question, il faudra se tourner vers les modèles des chapitres suivants.

CHAPITRE III

MÉTHODES STOCHASTIQUES

Au cours des dernières décennies, la venue de l'informatique a permis aux compagnies d'assurance de gérer une quantité d'information de plus en plus grande. Ces bases de données ont facilité l'intégration de modèles statistiques de plus en plus compliqués. Ceux-ci ont apporté une dimension nouvelle à l'analyse qui permet à la compagnie de mieux comprendre les données, et de mieux se préparer en fonction de sa santé financière et de sa stratégie d'affaires.

Dans cet ouvrage, on entend par *méthode stochastique* tout modèle qui suppose que la variable à modéliser possède un élément aléatoire. La plupart du temps, on lui associe une loi de probabilité, ou bien sinon, une mesure de son caractère variable, par exemple sa variance. Ainsi, on peut décrire le risque par une distribution ou bien par un intervalle de valeurs probables. Dans ce mémoire, on utilise l'expression *réserve* pour signifier l'espérance de la distribution associée à la variable aléatoire qui peut représenter la somme des pertes futures possibles.

Un des résultats les plus surprenants dans le domaine est que plusieurs modèles produisent exactement les mêmes réserves que la méthode Chain Ladder. En effet, le modèle de Poisson surdispersé (décrit ci-dessous) et le modèle de Mack (Mack, 1993) forment une liste non-exhaustive des modèles les plus connus. Cependant, en ajoutant

quelques hypothèses additionnelles, ces modèles permettent de déterminer le degré d'incertitude de la réserve, ce qui peut s'avérer une information capitale pour les stratégies financières de la compagnie. Ainsi, on peut voir la réserve comme n'étant que la pointe de l'iceberg, le risque étant représenté par le glacier au complet. Puisque la méthode Chain Ladder est déjà traditionnellement très utilisée, l'implantation de ces nouveaux modèles ne peut apporter que des avantages, puisque les réserves sont les mêmes.

3.1 Modèle de Poisson

Le modèle de Poisson est un modèle à caractère stochastique qui génère les mêmes réserves que la méthode Chain Ladder. Cette équivalence surprenante est un résultat néanmoins connu qui peut être démontré algébriquement (Wüthrich et Merz, 2008). Ce modèle suppose que les pertes incrémentales sont une réalisation d'une loi de probabilité. Cette méthode a été anticipée et développée par (Hachemeister et Stanard, 1975), (Mack, 1991) et (Renshaw et Verrall, 1998).

Afin de bien comprendre la théorie derrière l'implantation de ce modèle, plusieurs concepts reliés à la modélisation de variables aléatoires seront introduits au fur et à mesure dans cette section. Ceux-ci seront ensuite généralisés selon un modèle plus global dans la section 3.2.3.

3.1.1 Bases du modèle

Le modèle de Poisson suppose que la variable qui représente les pertes incrémentales $C_{i,j}$ suit une loi de Poisson, introduite dans la section 1.7.3, avec une moyenne $\lambda_{i,j}$ propre à chaque cellule (i, j) , tel que :

$$C_{i,j} \sim \text{Poisson}(\lambda_{i,j})$$

où $\lambda_{i,j}$ dépend de l'année d'accident i et la période de développement j . On suppose que les facteurs qui peuvent influencer sur la moyenne de cette cellule ne peuvent être que l'année d'accident et la période de développement. Cependant, lors de la construction du modèle, on peut aussi choisir d'autres facteurs explicatifs tel que l'année de calendrier, qui est une combinaison de i et de j . Néanmoins, dans ce mémoire, on utilisera seulement l'année d'accident i et de développement j et aucune autre combinaison supplémentaire. (Kaas, 2005) offre un exemple intéressant où les effets des années d'accident, de développement et de calendrier sont utilisés simultanément.

La façon la plus élémentaire de paramétriser $\lambda_{i,j}$ est de supposer que l'effet de l'année d'accident est indépendant de l'effet de la période de développement. De plus, on peut penser que pour chaque année d'accident, on a un portefeuille d'assurés différent, donc chaque année d'accident aura besoin d'un paramètre distinct pour représenter une réserve finale différente. Ensuite, dans la même logique, on suppose qu'à chaque période de développement, un pourcentage spécifique de la réserve finale sera réclamé, donc on aura besoin de paramètres différents selon la période de développement.

Il existe plusieurs types de paramétrisation, c'est-à-dire, de façons de décrire les facteurs explicatifs. Les moyennes prévues $\lambda_{i,j}$ ne sont pas affectées par le choix du type de paramétrisation. La paramétrisation la plus utilisée dans ce mémoire est la *paramétrisation avec un prédicteur additif et un lien multiplicatif*¹ définie ci-dessous. Afin d'alléger le texte, on l'appellera aussi : paramétrisation naturelle. On l'utilise pour son efficacité lors de l'optimisation des statistiques. Cette approche est inspirée de la paramétrisation *traitement*², largement utilisée en statistique lors de l'analyse de l'effet de divers traitements sur un groupe de base. Chaque paramètre représente alors la différence entre le groupe de base et le groupe ayant eu un traitement particulier. En ce

1. Log-link function and an additive predictor

2. Treatment

qui concerne les triangles de développement, nous utilisons :

$$\lambda_{i,j} = e^{\mu + \alpha_i + \beta_j} ,$$

où $i \in \{2, 3, \dots, I\}$ et $j \in \{2, 3, \dots, J\}$, qui est en fait un modèle linéaire généralisé avec fonction de lien multiplicatif et une loi de Poisson. En prenant le logarithme de l'équation, on peut reconnaître le système linéaire :

$$\log(\lambda_{i,j}) = \mu + \alpha_i + \beta_j ,$$

On observe que μ représente les pertes incrémentales moyennes pour la cellule de référence, qui est définie comme étant la première année d'accident et la première année de développement. Ensuite, α_i représente l'ajustement de la moyenne μ pour passer de la première ligne à la i^e ligne. De même, les paramètres β_j indiquent l'ajustement pour passer de la première colonne à la j^e colonne. Donc, par convention, on suppose que $\alpha_1 = \beta_1 = 0$. Ainsi, chaque cellule aura une combinaison unique qui la représente. Ici, pour un triangle de 9 x 11, un total de 19 paramètres sont utilisés, notés $\tilde{\mu}$:

$$\tilde{\mu} = (\mu, \alpha_2, \alpha_3, \dots, \alpha_9, \beta_1, \dots, \beta_{11}) .$$

Afin d'alléger la notation, on utilisera également :

$$\begin{aligned} \tilde{c} &= (c_{1,1}, c_{2,1}, \dots, c_{9,1}, c_{1,2}, \dots, c_{1,11}) \text{ et} \\ \tilde{C} &= (C_{1,1}, C_{2,1}, \dots, C_{9,1}, C_{1,2}, \dots, C_{1,11}) , \end{aligned}$$

qui représentent le vecteur des pertes incrémentales observées et le vecteur des variables aléatoires représentant les pertes incrémentales, respectivement.

3.1.2 Principe du maximum de vraisemblance

L'objectif est de déterminer les estimateurs des paramètres $\tilde{\mu}$ optimaux qui décrivent le mieux les pertes incrémentales observées \tilde{c} . Pour ce faire, on utilise le principe du *maximum de vraisemblance*. Celui-ci consiste à optimiser la statistique de vraisemblance L

qui représente la probabilité d'obtenir les données observées. Ainsi, on doit trouver les paramètres qui font en sorte que la probabilité d'obtenir les résultats observés soit maximale, d'où le nom *vraisemblance maximale*.

En théorie, il faut optimiser la fonction de probabilité conjointe des variables aléatoires représentant les pertes \tilde{C} , sachant les paramètres. En pratique, on suppose que les réalisations des variables $C_{i,j}$ sont indépendantes entre-elles lorsqu'on connaît les paramètres. Pour des fins de généralisation, notons $f(x)$ comme étant la fonction de probabilité de X , i.e. $\Pr(X = x)$ ou la fonction de densité de X . Ainsi, la fonction à optimiser est :

$$\begin{aligned} L(\tilde{\mu}|\tilde{c}) &= \prod_{i,j} f(\tilde{\mu} | c_{i,j}) \\ &= \prod_{i,j} \frac{e^{-\lambda_{i,j}} \lambda_{i,j}^{c_{i,j}}}{c_{i,j}!}, \end{aligned}$$

où $\lambda_{i,j} = e^{\mu + \alpha_i + \beta_j}$. Comme le produit d'un grand nombre de valeurs entre 0 et 1 rend la statistique très petite et difficile à traiter directement, il est courant d'optimiser la fonction de log-vraisemblance. Puisque c'est une transformation monotone, les paramètres optimaux sont les mêmes pour ces deux statistiques. La fonction de log-vraisemblance est donc :

$$\begin{aligned} l = \log(L(\tilde{\mu}|\tilde{c})) &= \sum_{i,j} \log(f(\tilde{\mu} | c_{i,j})) \\ &= \sum_{i,j} (-\lambda_{i,j} + c_{i,j} \log(\lambda_{i,j}) - \log(c_{i,j}!)) . \end{aligned} \quad (3.1)$$

Afin de simplifier la démonstration, supposons que le vecteur des paramètres $\tilde{\mu}$ possède p paramètres où $p = I + J - 1$. Aussi, afin d'alléger les démonstrations, soient les indicateurs $X_{(i,j),k}$ qui prennent les valeurs suivantes pour l'observation de la i^e ligne et de la j^e colonne :

$$X_{(i,j),k} = \begin{cases} 1 & , \text{ le } k^e \text{ paramètre de } \tilde{\mu} \text{ intervient dans le calcul de } \lambda_{i,j} \\ 0 & , \text{ autrement .} \end{cases}$$

On peut alors écrire :

$$\begin{aligned}\lambda_{i,j} &= e^{X_{(i,j),1}\mu + X_{(i,j),2}\alpha_2 + \dots + X_{(i,j),p}\beta_p} \\ &= e^{X_{i,j}\tilde{\mu}}.\end{aligned}\quad (3.2)$$

où $X_{i,j}$ est le vecteur des indicateurs $X_{(i,j),k}$. De plus, on peut aussi représenter les vecteurs pour toutes les observations à l'aide d'une matrice appelée *matrice catégorielle*. En voici un exemple lorsque les données sont enregistrées colonne par colonne, de haut en bas, de gauche à droite.

C	μ	α_2	α_3	α_4	α_5	α_6	α_7	α_8	α_9	β_2	β_3	β_4	β_5	β_6	β_7	β_8	β_9	β_{10}	β_{11}
17841110	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
19519117	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
19991172	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
19305646	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
18291478	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
18832520	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
17152710	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
16615059	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
16835453	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
7442433	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
6656520	1	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
6327483	1	0	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
5889791	1	0	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
5793282	1	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0
5741214	1	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0
5908286	1	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0
5111177	1	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0
5001897	1	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0
895413	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
941458	1	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
1100177	1	0	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
793020	1	0	0	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
689444	1	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0
581798	1	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0
524806	1	0	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0
553277	1	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0
489356	1	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0
407744	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
155395	1	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0

25833	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
321	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1

Par exemple, les pertes observées de la troisième ligne, deuxième colonne sont de

6327483. On voit que seules les colonnes vis-à-vis les paramètres α_3 et β_2 sont égales à 1.

Rappelons maintenant que l'objectif est de trouver les paramètres $\tilde{\mu}$ qui maximisent la fonction de vraisemblance. Alors, on veut que les premières dérivées soient égales à 0. Ainsi :

$$\begin{aligned} \frac{\partial l}{\partial \tilde{\mu}_k} &= \frac{\partial}{\partial \tilde{\mu}_k} \sum_{i,j} \log(\Pr(C_{i,j} | \tilde{\mu})) \\ &= \frac{\partial}{\partial \tilde{\mu}_k} \left(\sum_{i,j} (-\lambda_{i,j} + c_{i,j} \log(\lambda_{i,j}) - \log(c_{i,j}!)) \right) \\ &= \sum_{i,j} \left(-\frac{\partial}{\partial \tilde{\mu}_k} \lambda_{i,j} + c_{i,j} \frac{\partial}{\partial \tilde{\mu}_k} \log(\lambda_{i,j}) \right) = 0, \end{aligned}$$

avec :

$$\begin{aligned} \frac{\partial}{\partial \tilde{\mu}_k} \lambda_{i,j} &= \frac{\partial}{\partial \tilde{\mu}_k} e^{X_{(i,j),1}\mu + X_{(i,j),2}\alpha_2 + \dots + X_{(i,j),p}\beta_J} \\ &= e^{X_{(i,j),1}\mu + X_{(i,j),2}\alpha_2 + \dots + X_{(i,j),p}\beta_J} \\ &\quad \cdot \frac{\partial}{\partial \tilde{\mu}_k} (X_{(i,j),1}\mu + X_{(i,j),2}\alpha_2 + \dots + X_{(i,j),p}\beta_J) \\ &= X_{(i,j),k} e^{X_{(i,j),1}\mu + \dots + X_{(i,j),p}\beta_J} \\ &= X_{(i,j),k} \lambda_{i,j}, \end{aligned}$$

et :

$$\frac{\partial}{\partial \tilde{\mu}_k} \log(\lambda_{i,j}) = X_{(i,j),k}.$$

On obtient ainsi l'équation suivante :

$$\frac{\partial l}{\partial \tilde{\mu}_k} = \sum_{i,j} (-X_{(i,j),k} \lambda_{i,j} + c_{i,j} X_{(i,j),k}) \quad (3.3)$$

$$0 = \sum_{i,j} X_{(i,j),k} (c_{i,j} - \lambda_{i,j}). \quad (3.4)$$

où $\tilde{\mu}_k$ est le k^e paramètre de $\tilde{\mu}$. Cette démonstration est la même pour tous les $k \in \{1, \dots, p\}$. On obtient ainsi un système non-linéaire à p équations, p inconnues. Il est souvent difficile de trouver une solution entièrement analytique ou computationnellement simple à ce problème. C'est pourquoi, plusieurs algorithmes numériques ont

été mis au point pour profiter de la puissance de calcul brut des ordinateurs. Un algorithme fréquemment utilisé dans ce domaine est l'algorithme de Newton-Raphson. De plus amples détails sur son utilisation sont disponibles dans la section 3.1.6.

L'utilisation du principe du maximum de vraisemblance possède de nombreux avantages. En effet, en plus d'avoir d'excellentes propriétés théoriques, il est asymptotiquement sans-biais, convergent et possède la variance minimale parmi tous les estimateurs.

3.1.3 Système balancé

Le système d'équations (3.4) indique que le système est balancé. En d'autres mots, la somme des prévisions du modèle est égale à la somme des valeurs observées en fonction d'une condition, comme illustré ici :

$$\sum_{i,j} X_{(i,j),k} c_{i,j} = \sum_{i,j} X_{(i,j),k} \lambda_{i,j} .$$

où $X_{(i,j),k}$ représente la condition qui est fonction de la présence de certains paramètres ou non. Par exemple, dans une paramétrisation naturelle, lorsque $k = 1$, la matrice catégorielle se réfère au premier paramètre, μ . Comme ce paramètre apparaît pour *toutes* les observations, on en déduit que la somme des prévisions qui appartiennent au triangle doit égaler la somme des valeurs observées pour l'*ensemble* du triangle. Ensuite, considérons le deuxième paramètre α_2 . Ce paramètre n'est présent que lorsque les données appartiennent à la deuxième ligne. Ainsi, en suivant la même logique, la somme des prévisions de la deuxième ligne doit être égale à la somme des valeurs observées de la deuxième ligne. De même, les équations sous-jacentes pour les autres valeurs de k indiquent que la somme pour chaque ligne et chaque colonne doit être respectée.

Les tableaux 5.3 et 5.4 (p.130) en annexe illustrent numériquement le balance-

ment du modèle de Poisson. Finalement, il est important de mentionner que le modèle de Poisson n'est pas le seul qui donne un résultat balancé, et donc la solution d'un système balancé n'est pas unique (Mildenhall, 1999).

3.1.4 Restrictions du modèle

Tout d'abord, on suppose que les observations sont des réalisations de variables aléatoires indépendantes lorsque les paramètres sont fixés. Ensuite, le domaine positif de la loi de Poisson fait en sorte que le paramètre qui représente la moyenne λ doit être positif. Or, il se peut qu'en pratique, on observe plusieurs pertes incrémentales négatives. Et puisque le modèle balance pour toutes les lignes et colonnes, la somme des valeurs prévues doit égaler la somme des valeurs observées dans chaque cas. Donc, lorsque la somme de pertes est négative pour au moins une ligne ou colonne, alors le modèle ne peut être appliqué à moins d'effectuer des ajustements, soit au modèle, soit aux données.

Par ailleurs, la nature discrète de la loi fait en sorte que le domaine de la variable aléatoire qui représente les réserves le sera aussi. Plus de précisions sont disponibles dans la section 3.1.9.

3.1.5 Modélisation par unité d'exposition

La modélisation en tenant compte des *unités d'exposition*, notées $w_{i,j}$, est une caractéristique particulière des modèles stochastiques, qui leur offre beaucoup de flexibilité. Ceci nous permet d'accorder un degré d'importance différent à chaque observation. Dans la littérature des réserves stochastiques, pour tenir compte de l'exposition, on modélise les *pertes par unité de risque*, $Y_{i,j}$ avec la définition suivante :

$$Y_{i,j} = \frac{C_{i,j}}{w_{i,j}} .$$

Par exemple, (Wüthrich, 2003) utilise le volume d'affaire comme unité de risque. On peut alors interpréter les pertes par unité de risque comme des pertes moyennes par personne. Cet ajustement a pour but de tenir compte de la variabilité de la grandeur du portefeuille par année d'accident. Après avoir modélisé les pertes par unité de risque, on reconvertit les valeurs :

$$\hat{C}_{i,j} = w_{i,j} \hat{Y}_{i,j} .$$

Il faut comprendre que lorsque l'exposition est différente d'une observation à l'autre, le modèle est différent, même si on divise et ensuite on remultiplie par les mêmes valeurs. Donc l'espérance et l'erreur de prédiction de la réserve sera différente. Par ailleurs, il est à noter que la modélisation par unité de risque est une transformation invariante pour les opérations de multiplication par une constante. Donc, un modèle qui utilise $w_{i,j}$ comme exposition aura les mêmes réserves qu'un modèle avec $c \cdot w_{i,j}$ où c est une constante. Seuls comptent les rapports entre les facteurs d'exposition.

3.1.6 Algorithme d'optimisation numérique

Il existe plusieurs méthodes numériques pour résoudre un système d'équations non-linéaires. En effet, un logiciel tel que SAS possède déjà quelques routines pré-programmées telles que la *Trust Region Optimization* (Dennis Jr, Gay et Walsh, 1981), l'algorithme de Newton-Raphson, la *Double Dogleg Optimization* (Dennis et Mei, 1979), la *Conjugate Gradient Optimization* (Fletcher, 1987) et la *Nedler-Mead Simplex Optimization* (Olsson et Nelson, 1975).

À l'origine, la méthode de Newton-Raphson a été développée pour estimer la racine d'une fonction à une dimension, continue et dérivable. L'utilisateur donne la valeur de départ de l'algorithme, qui est, selon lui, la position de la racine. Ensuite, comme illustré dans la figure 3.1, l'algorithme trouve la tangente de la fonction estimée à ce point, longe la tangente pour trouver la racine de la droite engendrée par la tangente.

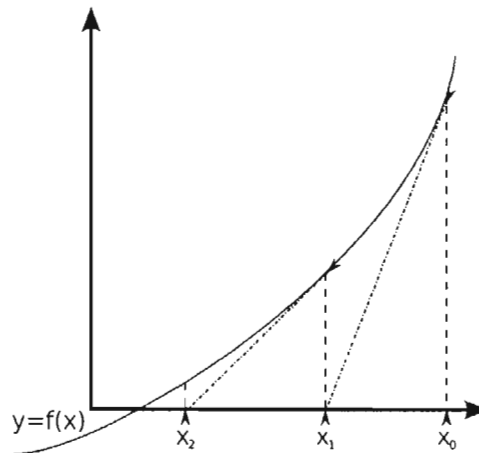


Figure 3.1 Itérations de l'algorithme Newton-Raphson

Si le nouveau point n'est pas assez proche de 0, on recommence la procédure avec ce nouveau point de départ.

Si la fonction fluctue beaucoup, alors elle peut difficilement être approximée par sa première dérivée, ce qui rend l'algorithme inefficace, et la probabilité qu'il y ait divergence augmente. Toutefois, la fonction de vraisemblance est généralement une fonction aplatie, qui ne fluctue pas beaucoup pour la grande majorité des lois présentées dans ce mémoire.

Par ailleurs, puisqu'il faut estimer plusieurs paramètres à la fois, on utilise la version multivariée de l'algorithme de Newton-Raphson. Pour cela, on évalue les tendances linéaires de chaque variable, mais aussi les tendances linéaires entre chaque combinaison de variable. Toute cette information est placée dans une matrice qu'on appelle la matrice Hessienne, notée H . Rappelons qu'on cherche les tendances linéaires *des premières dérivées* de la fonction de vraisemblance. Nous aurons donc besoin des deuxièmes dérivées de la fonction de log-vraisemblance de l'équation (3.1), ou bien des

premières dérivées de l'équation (3.3), qui représentent la matrice H. Donc, à partir de l'équation (3.3), les valeurs recherchées de la matrice Hessienne sont :

$$\begin{aligned}
 \frac{\partial l}{\partial \tilde{\mu}_k} &= \sum_{i,j} X_{(i,j),k} (c_{i,j} - \lambda_{i,j}) \quad , \\
 \frac{\partial^2 l}{\partial \tilde{\mu}_k \partial \tilde{\mu}_m} &= \frac{\partial}{\partial \tilde{\mu}_m} \sum_{i,j} X_{(i,j),k} (c_{i,j} - \lambda_{i,j}) \\
 &= \sum_{i,j} X_{(i,j),k} \frac{\partial}{\partial \tilde{\mu}_m} (c_{i,j} - \lambda_{i,j}) \\
 &= \sum_{i,j} -X_{(i,j),k} \frac{\partial}{\partial \tilde{\mu}_m} \lambda_{i,j} \\
 &= \sum_{i,j} -X_{(i,j),k} X_{(i,j),m} \lambda_{i,j} \quad .
 \end{aligned} \tag{3.5}$$

Pour un vecteur de paramètres donné $\tilde{\mu}$, les valeurs calculées de l'équation (3.5) sont ensuite placées de la façon suivante :

$$H(\tilde{\mu}) = \begin{pmatrix} \frac{\partial^2 l}{\partial \tilde{\mu}_1 \partial \tilde{\mu}_1} & \frac{\partial^2 l}{\partial \tilde{\mu}_1 \partial \tilde{\mu}_2} & \cdots & \frac{\partial^2 l}{\partial \tilde{\mu}_1 \partial \tilde{\mu}_p} \\ \frac{\partial^2 l}{\partial \tilde{\mu}_2 \partial \tilde{\mu}_1} & \frac{\partial^2 l}{\partial \tilde{\mu}_2 \partial \tilde{\mu}_2} & & \\ \vdots & & \ddots & \vdots \\ \frac{\partial^2 l}{\partial \tilde{\mu}_p \partial \tilde{\mu}_1} & & \cdots & \frac{\partial^2 l}{\partial \tilde{\mu}_p \partial \tilde{\mu}_p} \end{pmatrix} ,$$

où $H(\tilde{\mu})$ est la matrice hessienne évaluée au vecteur de paramètres $\tilde{\mu}$.

Ainsi, pour résoudre le système d'équations, l'utilisateur choisit des valeurs de départ pour les paramètres, et ensuite, il évalue si les équations s'approchent de 0. Soit $U(\tilde{\mu})$ le système d'équations des premières dérivées de l'équation (3.1), évaluées avec le vecteur de paramètres $\tilde{\mu}$:

$$U(\tilde{\mu}) = \begin{pmatrix} U(\tilde{\mu}_1) \\ U(\tilde{\mu}_2) \\ \vdots \\ U(\tilde{\mu}_p) \end{pmatrix} = \begin{pmatrix} \sum_{i,j} X_{(i,j),1} (c_{i,j} - \lambda_{i,j}) \\ \sum_{i,j} X_{(i,j),2} (c_{i,j} - \lambda_{i,j}) \\ \vdots \\ \sum_{i,j} X_{(i,j),p} (c_{i,j} - \lambda_{i,j}) \end{pmatrix}.$$

Rappelons aussi que

$$\lambda_{i,j} = e^{X_{i,j}\tilde{\mu}}.$$

Lorsque $\tilde{\mu}$ change, alors λ change et donc finalement $U(\tilde{\mu})$ fluctuera en conséquence. Rappelons que la fonction de log-vraisemblance (3.1) est optimisée lorsque $U(\tilde{\mu}) = 0$. On évalue $U(\tilde{\mu}^{(0)})$ où $\tilde{\mu}^{(0)}$ représente l'ensemble des paramètres qu'on fixe à une valeur de départ, et où $\tilde{\mu}_k$ indique les estimés des paramètres à la k^e itération. Le nouveau vecteur de paramètres se trouve à l'aide de la formule suivante :

$$U(\tilde{\mu}^{(k+1)}) = \tilde{\mu}^{(k)} - H^{-1}(\tilde{\mu}^{(k)})U(\tilde{\mu}^{(k)}).$$

où $H^{-1}(\tilde{\mu}^{(k)})$ est l'inverse de la matrice hessienne évaluée à $\tilde{\mu}^{(k)}$.

Finalement, lorsque l'algorithme converge, il faut vérifier les paramètres pour détecter les aberrations. Lorsque certains paramètres tendent vers l'infini, il faut essayer d'autres points de départs et essayer de déterminer si la convergence se fait toujours vers le même point.

3.1.7 Autres paramétrisations

Comme mentionné dans la section 3.1.1, il existe plusieurs façons de décrire l'influence des facteurs explicatifs sur la moyenne des pertes incrémentales. Jusqu'ici, la paramétrisation naturelle a été utilisée et, pour le triangle 9 x 11, nous avons eu besoin de 19 paramètres.

Bien entendu, il existe d'autres types de paramétrisation, l'une étant la *paramétrisation intuitive*³. L'idée est d'associer directement un paramètre pour chaque ligne et chaque colonne, sans qu'il y ait de cellule de référence. Ainsi,

$$\lambda_{i,j} = \alpha_i \beta_j ,$$

où α_i représente le paramètre associé à l'année d'accident i et représente l'encours cumulé ultime, alors que β_j est le paramètre associé à la période de développement j et représente la portion de l'encours cumulé ultime réclamé à chaque période. On obtient alors la matrice catégorielle suivante :

C	α_1	α_2	α_3	α_4	α_5	α_6	α_7	α_8	α_9	β_1	β_2	β_3	β_4	β_5	β_6	β_7	β_8	β_9	β_{10}	β_{11}
17841110	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
19519117	0	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
19991172	0	0	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
19305646	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
18291478	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
18832520	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
17152710	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0
16615059	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0
16835453	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0
7442433	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
6656520	0	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
6327483	0	0	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
5889791	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
5793282	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
5741214	0	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0
5908286	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0
5111177	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0
5001897	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0
895413	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
941458	0	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
1100177	0	0	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
25833	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0
321	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1

Cette approche possède un trop grand nombre de paramètres. En effet, nous avons 9 lignes et 11 colonnes, donc 20 paramètres en tout. L'optimisation n'est possible que si

3. inspirée de la paramétrisation *somme* (Sum) utilisée en statistique

on ajoute un nombre suffisant de contraintes. Dans ce cas-ci, on ajoute la contrainte :

$$\sum_{j=1}^J \beta_j = 1 .$$

Intuitivement, cette contrainte indique que les β_j doivent être des proportions. Il s'avère que ce système peut être résolu directement, cependant, l'algorithme converge moins facilement vers la solution. Néanmoins, la réserve espérée, ainsi que la provision reste la même, peu importe la paramétrisation. D'autre part, on peut résoudre un système d'équations élémentaire pour passer d'une paramétrisation à l'autre tant au niveau de l'espérance des paramètres que de leur écart-type. Bref, il est suggéré d'utiliser la paramétrisation naturelle tout en gardant à l'esprit qu'on peut toujours convertir en une nouvelle paramétrisation si cela s'avère nécessaire.

3.1.8 Courbe d'Hoerl

La flexibilité des modèles stochastiques permet d'incorporer des tendances dans le modèle. Par exemple, au lieu d'estimer les paramètres de chaque période de développement individuellement, on peut supposer qu'il y a un effet linéaire et log-linéaire dans le temps. Cette façon de modéliser est appelée la *courbe de Hoerl* (England et Verrall, 2002). Notons que puisqu'il y a deux tendances à estimer, seuls deux paramètres seront nécessaires, contrairement aux 10 paramètres nécessaires à une estimation par colonne. Mathématiquement, on représente le modèle de la façon suivante :

$$\beta_j = \gamma_1 \log(j) + \gamma_2 j .$$

Les paramètres représentant les années d'accident α_i restent intacts, alors que les paramètres associés aux périodes de développement β_j sont remplacés par γ_1 et γ_2 . Cette méthode est utilisée lorsqu'on pense que le développement sur les colonnes peut être modélisé par une fonction simple. De plus, cette méthode possède l'avantage de pouvoir estimer les réserves entre les fins de périodes, et aussi de pouvoir extrapoler après la dernière période observée. Les paramètres estimés selon un modèle de Poisson avec une

courbe d'Hoerl, ainsi que les réserves obtenues sont disponibles dans le tableau 5.6 (p. 133) en annexe.

3.1.9 Paramètre de surdispersion

La plupart des modèles présentés dans la section 1.7 possèdent deux paramètres. Cette flexibilité naturelle permet de capter aussi bien la moyenne que la variance des données. Il a été démontré que certaines de ces lois ont une variance qui est fonction de la moyenne. Cependant, certaines distributions, tel que la loi de Poisson, ne possèdent pas suffisamment de paramètres pour modéliser aussi bien la variance que l'espérance puisque l'espérance est égale à la variance. On voudrait trouver une façon de rendre ces modèles plus flexibles. L'idée est d'ajouter artificiellement un paramètre supplémentaire pour forcer la loi à avoir une variance qui reflète mieux les données. Ainsi, nous définissons un *paramètre de surdispersion*⁴, noté ϕ , comme étant le facteur d'ajustement de la variance, lorsque la variance de la loi est fonction de son espérance.

Dans le cas de la loi gamma, nous n'avons pas besoin d'ajouter artificiellement de nouveaux paramètres car la loi possède déjà deux paramètres. Ainsi, le paramètre de surdispersion est introduit suite à une reparmétrisation de sorte à ce qu'on isole l'effet de la moyenne $\lambda_{i,j}$ et de la surdispersion ϕ , comme démontré dans la section 1.7.4. Nous obtenons alors un modèle où :

$$E[C_{i,j}] = \lambda_{i,j} , \text{ Var}[C_{i,j}] = \phi \lambda_{i,j}^2 ,$$

et donc ϕ complète une relation où la variance est fonction du carré de l'espérance.

En ce qui concerne la loi normale, le paramètre de surdispersion ϕ est exactement le paramètre qui représente la variance de la loi, comme démontré dans les équations

4. Dispersion Parameter

(3.6) et (3.7) :

$$C_{i,j} \sim N(\lambda_{i,j} ; \sigma^2) , \quad (3.6)$$

$$\sigma^2 = \phi , \quad (3.7)$$

$$C_{i,j} \sim N(\lambda_{i,j} ; \phi) .$$

Finalement, pour un modèle de Poisson, l'introduction d'un facteur de surdispersion doit se faire en utilisant le principe de la quasi-vraisemblance⁵. Pour être bref, ce principe met l'accent sur la relation entre l'espérance et la variance de la loi plutôt que sur la fonction de densité et son domaine. En d'autres mots, il suffit de trouver une statistique qui engendre la relation entre l'espérance et la variance désirée sur un domaine préétabli. Dans le cas de la loi de Poisson, le nouveau modèle est appelé loi de *Poisson surdispersée*⁶ et peut être défini à l'aide du raisonnement ci-dessous.

Définissons $X_{i,j}$ une variable aléatoire qui suit une loi de Poisson de moyenne et de variance toutes deux égales à $\lambda_{i,j}$:

$$X_{i,j} \sim \text{Poisson}(\lambda_{i,j}) ,$$

et aussi, soit :

$$\begin{aligned} Y_{i,j} &\sim \text{Poisson}\left(\frac{\lambda_{i,j}}{\phi}\right) , \quad Y_{i,j} \in \{0, 1, \dots\} , \\ Z_{i,j} &= \phi Y_{i,j} , \quad Z_{i,j} \in \{0, \phi, 2\phi, \dots\} , \end{aligned}$$

où ϕ est le facteur de surdispersion désiré. Nous avons alors :

$$\begin{aligned} E[Y_{i,j}] &= \frac{\lambda_{i,j}}{\phi} , \quad \text{Var}[Y_{i,j}] = \frac{\lambda_{i,j}}{\phi} , \\ E[Z_{i,j}] &= \lambda_{i,j} , \quad \text{Var}[Z_{i,j}] = \phi \lambda_{i,j} . \end{aligned}$$

5. Quasi-likelihood

6. Overdispersed Poisson Model

La variable aléatoire $Z_{i,j}$ possède alors une relation entre l'espérance et la variance. De plus, elle se comporte comme une loi de Poisson, mais qui est définie seulement sur les multiples du facteur de surdispersion, ce qui fait en sorte que pour un tel modèle, la distribution des réserves sera en escaliers (Wüthrich et Merz, 2008).

Le paramètre de surdispersion peut être estimé par au moins trois façons : le principe de vraisemblance, le principe des résidus de la déviance, le principe des résidus de Pearson.

Principe de vraisemblance

Lorsque la variable de surdispersion ϕ est présente dans l'équation de quasi-vraisemblance, alors il suffit de l'obtenir en même temps que les variables relatives aux moyennes lorsqu'on optimise la vraisemblance. Certains logiciels d'optimisation numérique, tel que SAS, procèdent automatiquement à ce type d'optimisation. De plus, on peut avoir une mesure du degré de l'incertitude de ce paramètre grâce à la matrice hessienne.

Principe de la déviance

L'idée est de trouver une mesure de dispersion entre les observations et les moyennes. Soit D la mesure de la déviance définie comme suit :

$$D = 2(l(\lambda_1, \dots, \lambda_N) - l(c_1, \dots, c_N)) ,$$

où N est le nombre d'observations dans le triangle. C'est un test de rapport de vraisemblance entre la log-vraisemblance du modèle saturé $l(c_1, \dots, c_N)$, qui suppose que toutes les moyennes sont égales aux observations, et la log-vraisemblance du modèle réduit $l(\lambda_1, \dots, \lambda_N)$, qui suppose que les moyennes sont une fonction des paramètres $\tilde{\mu}$ seulement. Alors, on estime le facteur de surdispersion issu du principe de la déviance,

noté ϕ_D , par :

$$\widehat{\phi_D} = \frac{D}{N-Q} = \frac{2(l(\lambda_1, \dots, \lambda_N) - l(c_1, \dots, c_N))}{N-Q} \quad (3.8)$$

où N est le nombre d'observations et Q le nombre de paramètres qui estiment les moyennes. Si le nombre d'observations est grand, on peut montrer que la déviance suit approximativement une loi Chi-deux (Nelder et Wedderburn, 1989), comme illustré ici :

$$D \sim \chi^2_{N-Q}.$$

où χ^2_{N-Q} est une loi Chi-deux avec $N-Q$ degrés de liberté. Ainsi, le paramètre de surdispersion peut être interprété intuitivement comme étant le paramètre d'échelle d'une loi Chi-deux.

Principe de Pearson

L'idée est la même que pour le principe de la déviance, sauf qu'ici, on calcule le carré de la différence entre les observations et les moyennes, ajustés selon la variance théorique du modèle. On suppose ensuite que cette statistique suit une loi Chi-deux, comme proposé ici :

$$\widehat{\phi_P} = \frac{\chi^2}{N-Q},$$

où χ^2 est la statistique de Pearson :

$$\chi^2 = \sum_{i,j} \left(\frac{(c_{i,j} - \lambda_{i,j})^2}{\text{Var}(C_{i,j})} \right). \quad (3.9)$$

Dans tous les cas, le paramètre de surdispersion n'a qu'un effet marginal sur l'estimation des moyennes. À l'inverse, son influence est souvent non-négligeable sur la variance des paramètres. En conséquence, le paramètre de surdispersion a un impact souvent important sur la variance de la réserve, en plus de receler de l'information pertinente sur la nature du risque. De plus, puisqu'il existe une indépendance entre ϕ et $\tilde{\mu}$

au niveau de l'optimisation, on estime en général le paramètre de surdispersion qu'après avoir calculé les paramètres $\tilde{\mu}$ comme en témoignent les équations (3.8) et (3.9). D'autres détails sont disponibles dans la section 3.2.3.

Le calcul du facteur de surdispersion selon l'un ou l'autre de ces principes peut engendrer des valeurs sensiblement différentes. Le chapitre 4 fait une étude approfondie de l'impact du facteur de surdispersion selon le principe choisi.

3.1.10 Variance des estimateurs des paramètres

Les estimateurs trouvés avec la statistique de vraisemblance ne sont qu'un estimé de la valeur exacte des *vrais* paramètres. En d'autres mots, le fait qu'on procède à une estimation indique qu'il devrait y avoir une incertitude associée aux paramètres. Si l'incertitude est grande, alors l'intervalle qui contient les valeurs plausibles est souvent plus grand. Ainsi, il arrive fréquemment d'associer une variance et parfois même une distribution aux paramètres estimés.

Le principe de vraisemblance maximale permet d'établir une mesure de la variance des paramètres. En effet, si la loi choisie possède certaines conditions de régularité (comme c'est le cas pour tous les modèles traités dans cet ouvrage), alors les estimateurs suivent asymptotiquement une loi normale multivariée (Klugman et al., 1998), comme présenté ici :

$$\hat{\mu} \sim N\left(\tilde{\mu}; \widehat{\Sigma}_{\tilde{\mu}}^2\right),$$

où $\tilde{\mu}$ est le *vrai* paramètre, $\hat{\mu}$ est l'estimé du maximum de vraisemblance et la matrice de variance-covariance $\widehat{\Sigma}_{\tilde{\mu}}^2$ est l'inverse de la matrice hessienne. Afin de mieux comprendre en quoi l'inverse de la matrice hessienne est une bonne indication de la variance des paramètres, observons la situation à une dimension suivante décrite par la figure 3.2.



Figure 3.2 Le paramètre peut varier plus lorsque la courbe est aplatie

La courbe représente la fonction de vraisemblance. Lorsque la fonction de vraisemblance varie beaucoup au point optimal, la deuxième dérivée évaluée à ce point est très grande. Ceci indique que si on choisit un autre paramètre, la fonction de vraisemblance décroît rapidement, alors il n'y a pas un grand intervalle de valeurs plausibles autour du point optimal. Par le fait même, on conclut que la variance du paramètre est petite. À l'inverse, si la fonction est aplatie, alors le raisonnement contraire est utilisé et on conclut à une grande variance. Donc, la variance est inversement reliée à la matrice hessienne.

Grâce à la variance des estimateurs, il devient possible de déterminer si un estimateur est significativement différent d'un autre ou non (test de Wald, test du rapport de vraisemblance, etc.), ce qui permet de mieux calibrer le modèle ou d'en proposer un qui en contient moins.

Par ailleurs, il est important de mentionner que la matrice de variance-covariance doit être ajustée si le principe de quasi-vraisemblance maximale est utilisé. En effet, l'introduction du paramètre de surdispersion a un effet sur la variance des paramètres (alors qu'il n'affecte pas directement leur espérance). Ainsi, plus le paramètre est élevé, plus la variance le sera aussi. En conséquent, on obtient la variance-covariance finale en multipliant le facteur de surdispersion par la matrice de variance-covariance originelle.

ment obtenue lors de l'estimation qui ne tient pas compte du paramètre de surdispersion.

3.1.11 Variance de la réserve

La grande force des modèles stochastiques est qu'on peut trouver le degré d'incertitude de la réserve. Nous appelons *variance de la réserve*, une mesure de la fluctuation de la réserve qui peut être causée par l'ensemble des types d'erreur d'assurance. On définit l'*erreur* comme étant la variabilité ou la fluctuation associée à un élément du modèle. (Taylor, 1987) indique que l'erreur d'assurance totale se décompose en quatre parties principales⁷ :

1. l'erreur statistique ,
2. l'erreur d'estimation ,
3. l'erreur de sélection ,
4. l'erreur de spécification .

L'erreur statistique est le bruit causé par la distribution choisie. Par exemple, pour un individu donné, la perte suit une loi quelconque qui aura des réalisations possiblement différentes à chaque observation. Ce type d'erreur est généralement inévitable et présent dans presque tous les modèles.

L'erreur d'estimation indique l'incertitude des paramètres optimaux estimés par rapport à ce que pourraient être les vrais paramètres. Par exemple, lorsqu'on utilise les estimateurs issus du principe de vraisemblance maximale, on suppose la normalité asymptotique.

7. 1. Statistical Error (Process Error) 2. Estimation Error 3. Selection Error 4. Specification Error

L'erreur de sélection spécifie l'incertitude face à des variables cachées qui ne sont pas présentes dans le modèle, mais qui en réalité ont un impact sur le modèle. Un bon exemple est lorsqu'il y a dans la réalité un effet calendrier, mais qu'on ne l'inclut pas dans le modèle.

L'erreur de spécification essaie d'identifier la pertinence de la loi choisie par rapport à d'autres lois. Par exemple, l'incertitude face au paramètre p du modèle de Tweedie (section 3.2.3) sert à déterminer lequel parmi le modèle de Poisson (section 3.2.1), gamma (section 3.2.2), ou une combinaison des deux (section 3.2.3) est le plus approprié. Ce type d'erreur n'est pas tenu en compte dans aucun modèle présenté dans ce mémoire.

Finalement, il est à noter que ces quatre types d'erreurs ne constituent qu'une partie de l'incertitude *d'assurance*. Lorsqu'on établit une réserve, il existe d'autres risques à considérer également. Par exemple, les risques financiers (rendement des actifs dans la réserve, taux d'intérêt, inflation sociale), les risque de souscription (risque de défaut, risque moral), les risques juridiques (changement dans les lois qui gouvernent l'assurance), environnementaux (catastrophes naturelles), les risques du marché (offre et demande, marketing, stratégies des concurrents), etc. De nouveaux types d'analyses émergent pour tenir compte de ces types de risques, notamment l'analyse financière dynamique⁸, et la gestion du risque de l'entreprise⁹, mais ces concepts vont au-delà des objectifs de ce mémoire et ne seront pas considérés.

Une statistique courante reliée à la variance de la réserve R_i est l'*erreur quadratique moyenne de prédiction* (EQMP)¹⁰. Cette statistique est une combinaison de

8. Dynamic Financial Analysis

9. Enterprise Risk Management

10. Mean Square Error of Prediction

l'erreur statistique et de l'erreur d'estimation. Soit $\hat{C}_{i,j}$, la variable aléatoire représentant les pertes incrémentales de la cellule (i, j) lorsque les vrais paramètres sont inconnus et ont été estimés, alors que $C_{i,j}$ est la même variable aléatoire lorsque les paramètres sont *vrais*, on définit l'EQMP de la façon suivante (England et Verrall, 2002) :

$$\begin{aligned} \text{EQMP}[\hat{C}_{i,j}] &= E[(C_{i,j} - \hat{C}_{i,j})^2] \\ &= E[((C_{i,j} - E[C_{i,j}]) - (\hat{C}_{i,j} - E[\hat{C}_{i,j}]))^2] \\ &= E[(C_{i,j} - E[C_{i,j}])^2] - 2E[(C_{i,j} - E[C_{i,j}])(\hat{C}_{i,j} - E[\hat{C}_{i,j}])] \\ &\quad + E[(\hat{C}_{i,j} - E[\hat{C}_{i,j}])^2] . \end{aligned}$$

En supposant que $E[(C_{i,j} - E[C_{i,j}])(\hat{C}_{i,j} - E[\hat{C}_{i,j}])] = 0$, ou en d'autres mots, l'erreur d'estimation est indépendante de l'erreur statistique, il ne reste que :

$$\begin{aligned} \text{EQMP}[\hat{C}_{i,j}] &\approx E[(C_{i,j} - E[C_{i,j}])^2] + E[(\hat{C}_{i,j} - E[\hat{C}_{i,j}])^2] \\ &= \text{Var}[C_{i,j}] + \text{Var}[\hat{C}_{i,j}] , \end{aligned}$$

qui peut être interprété comme l'erreur statistique $\text{Var}[C_{i,j}]$ plus l'erreur d'estimation $\text{Var}[\hat{C}_{i,j}]$.

Pour un modèle de Poisson surdispersé, l'erreur statistique est $\text{Var}[C_{i,j}] = \phi\lambda_{i,j}$, alors que l'erreur d'estimation est approximée linéairement par la méthode delta (Klugman et al., 1998) tel que proposé dans les équations suivantes :

$$E[\hat{C}_{i,j}] = \lambda_{i,j} = e^{\eta_{i,j}} ,$$

où $\eta_{i,j} = \sum_{k=1}^p X_{(i,j),k} \tilde{\mu}_k$ et donc :

$$\begin{aligned} \text{Var}[\hat{C}_{i,j}] &= \frac{\partial \lambda_{i,j}}{\partial \eta_{i,j}} \text{Var}[\eta_{i,j}] \frac{\partial \lambda_{i,j}}{\partial \eta_{i,j}} \\ &= \lambda_{i,j}^2 \text{Var}[\eta_{i,j}] , \end{aligned}$$

et :

$$\text{EQMP}[\hat{C}_{i,j}] = \phi\lambda_{i,j} + \lambda_{i,j}^2 \text{Var}[\eta_{i,j}] .$$

De même, si on souhaite calculer l'EQMP de la réserve $\text{EQMP}[\hat{R}_i]$ pour une année d'accident i , il faut additionner l'erreur statistique et d'estimation de toutes les pertes incrémentales futures. En ce qui concerne l'erreur statistique, on suppose que lorsque les paramètres $\tilde{\mu}$ sont connus et fixes, l'erreur statistique est indépendante pour chaque observation. Par contre, puisque certains paramètres interviennent souvent dans plusieurs pertes incrémentales futures, il faut tenir compte des corrélations entre ces observations prévues. Pour ce faire, définissons L comme étant l'ensemble qui représente les périodes de développement futures pour l'année d'accident i . Nous avons alors :

$$\begin{aligned} \text{EQMP}[\hat{R}_i] = \text{EQMP}_L[\hat{C}_{i,j}] &= \sum_{(i,j) \in L} \phi\lambda_{i,j} + \sum_{(i,j) \in L} \lambda_{i,j} \text{Var}[\eta_{i,j}] \\ &+ \sum_{(i,j) \in L, j_1 \neq j_2} \lambda_{i,j_1} \lambda_{i,j_2} \text{Cov}(\eta_{i,j_1}, \eta_{i,j_2}) . \end{aligned}$$

Voici un exemple pour calculer $\text{EQMP}[\hat{R}_3]$, où il reste 2 paiements incrémentaux à déterminer. On utilise la paramétrisation naturelle, $i = 3$ et L représente les périodes de développement restantes, c'est-à-dire 10 et 11. Ainsi :

$$\begin{aligned} \text{EQMP}[\hat{R}_3] &= \phi\lambda_{3,10} + \phi\lambda_{3,11} \\ &+ \lambda_{3,10}^2 \text{Var}[\eta_{3,10}] + \lambda_{3,11}^2 \text{Var}[\eta_{3,11}] \\ &+ \lambda_{3,10}\lambda_{3,11} \text{Cov}(\eta_{3,10}, \eta_{3,11}) , \end{aligned}$$

et :

$$\begin{aligned} \text{Cov}(\eta_{3,10}, \eta_{3,11}) &= \text{Cov}(\mu + \alpha_3 + \beta_{10}, \mu + \alpha_3 + \beta_{11}) \\ &= \text{Cov}(\mu, \mu) + \text{Cov}(\mu, \alpha_3) + \text{Cov}(\mu, \beta_{11}) \\ &+ \text{Cov}(\alpha_3, \mu) + \text{Cov}(\alpha_3, \alpha_3) + \text{Cov}(\alpha_3, \beta_{11}) \\ &+ \text{Cov}(\beta_{10}, \mu) + \text{Cov}(\beta_{10}, \alpha_3) + \text{Cov}(\beta_{10}, \beta_{11}) . \end{aligned}$$

On trouve les covariances entre deux paramètres dans la matrice de variance-covariance, ajustée par le facteur de surdispersion.

Finalement, pour calculer l'EQMP de la réserve totale ($\text{EQMP}[\hat{R}]$), on applique le même principe, mais on ajoute les interactions entre-lignes aussi. Soit Δ , l'ensemble qui représente les pertes incrémentales futures pour toutes les années d'accident, nous avons alors :

$$\begin{aligned} \text{EQMP}[\hat{R}] &= \sum_{(i,j) \in \Delta} \phi \lambda_{i,j} \\ &+ \sum_{(i,j) \in \Delta} \lambda_{i,j}^2 \text{Var}[\eta_{i,j}] \\ &+ \sum_{(i,j) \in \Delta, (i_1,j_1) \neq (i_2,j_2)} \lambda_{i_1,j_1} \lambda_{i_2,j_2} \text{Cov}(\eta_{i_1,j_1}, \eta_{i_2,j_2}) . \end{aligned}$$

3.1.12 Approche bayésienne

Les modèles bayésiens sont issus d'une école de pensée qui a longtemps été considérée comme une alternative au mode de pensée fréquentiste. Depuis peu, pour certaines situations, les deux théories peuvent proposer des résultats identiques (même si leur interprétation est fondamentalement différente), rendant le cas fréquentiste un cas particulier des modèles bayésiens pour certains contextes précis. La caractéristique principale des modèles bayésiens est leur habileté à incorporer de l'information *a priori*. Lorsqu'on utilise des lois informatives, cela peut être une source de préoccupation, car on peut penser que ça enlève de l'objectivité au modèle. Néanmoins, il est important d'avoir des outils qui nous permettent d'ajuster le modèle en fonction de notre jugement car les situations où un ajustement extérieur au modèle est nécessaire sont très courantes.

Habituellement, lors de l'analyse des réserves, un modèle fréquentiste est utilisé, après quoi les valeurs qui semblent incorrectes sont modifiées artificiellement.

Jusqu'à aujourd'hui, en pratique, lorsque le modèle doit être ajusté, on emploie un modèle fréquentiste, après quoi on modifie artificiellement les valeurs qui semblent incorrectes. Cette façon de procéder est peu compatible avec la théorie statistique. Les modèles bayésiens permettent d'incorporer de la connaissance *a priori* dans le modèle dans un contexte statistique, tout en restant mathématiquement rigoureux.

Dans ce mémoire, une connaissance *a priori* signifie une connaissance qu'on possède sur certains paramètres du modèle avant que les observations soient incorporées dans celui-ci. Cette connaissance peut être parfaite : on sait que certains paramètres prennent une valeur précise sans aucune ambiguïté. Elle peut être partielle : on pense connaître la valeur approximative de certains paramètres avec un certain degré de précision. Finalement, la connaissance peut être nulle : on n'a aucune idée sur la valeur des paramètres. Notons que dans le dernier cas, on se base uniquement sur le modèle pour nous fournir les meilleurs estimés des paramètres. Dans certains cas, une analyse avec une loi *a priori* non-informative peut nous ramener à une analyse fréquentiste.

Il existe plusieurs types d'information *a priori*. L'idée est de transformer l'information brute *a priori* en une information sur des paramètres du modèle. Par exemple, on peut supposer que les pertes ultimes sont connues, ce qui est équivalent à utiliser la méthode Bornhuetter-Ferguson (section 2.3). Celle-ci peut être traduite en modèle bayésien de la façon suivante : on suppose que les pertes incrémentales suivent une loi prédéterminée et que les paramètres de cette loi dépendent entre autres des pertes ultimes, qui sont traitées comme un paramètre connu.

Si on pense que les changements ont un très grand impact sur la valeur des facteurs de développement, alors on utilise une loi *a priori* avec une petite variance autour des

estimés obtenus initialement. Sinon, on augmente la variance, rendant ainsi le modèle de plus en plus près du modèle fréquentiste. Donc, le modèle bayésien permet de jouer avec le degré de certitude qu'on a envers le changement observé.

Pour implanter un modèle bayésien, on utilise généralement un algorithme de simulation. De plus amples détails sont disponibles dans la section 3.1.13.

3.1.13 Générer la distribution de la réserve

Nous avons vu que l'erreur quadratique moyenne de prédiction (section 3.1.11) est une mesure du degré d'incertitude de la réserve qui peut s'avérer utile. Cependant, on peut aussi être intéressé à connaître la distribution de la réserve, ce qui nous donne une idée beaucoup plus précise de la nature du risque. La complexité du modèle quand il possède plusieurs types d'erreurs (section 3.1.11) et les interactions entre les variables font en sorte que distribution de la réserve est souvent difficile à être exprimée analytiquement. C'est pourquoi on emploie des techniques numériques basées sur la puissance computationnelle brute de l'ordinateur. Il existe trois approches principales pour générer la distribution de la réserve.

I - Méthode avec les moments

L'approche proposée dans (Mack, 1993) consiste à construire un modèle où on est en mesure de trouver analytiquement les deux premiers moments de la réserve. Ensuite, on suppose que la réserve suit une loi normale ou log-normale avec des paramètres qui correspondent à ces moments. Naturellement, cette supposition n'est qu'une approximation de la réalité, cependant, l'approche est intuitive, facilement implantable et constitue un bon modèle de référence.

II - Bootstrap

Cette approche a été initialement développée par (Efron et Tibshirani, 1993). Elle a ensuite été adaptée aux réserves actuarielles par (England et Verrall, 1999). Brièvement, la procédure consiste à trouver les résidus des observations par rapport à un modèle choisi, créer une pseudo-réalité en échantillonnant avec remise à partir de ces résidus, et calculer la réserve de cette pseudo-réalité. En répétant ce processus un grand nombre de fois, on obtient une distribution de la réserve. De plus, lorsqu'on utilise un modèle qui produit les mêmes réserves que la méthode Chain Ladder, par exemple le modèle de Poisson surdispersé (section 3.2.1), ou le modèle de Mack (Mack, 1999), cet algorithme peut facilement être implanté sans avoir besoin d'un logiciel statistique.

III - Monte Carlo par chaînes de Markov

Finalement, on peut utiliser la simulation de Monte-Carlo. Puisqu'on suppose que les pertes incrémentales suivent une distribution connue, alors on peut simuler un grand nombre de réalisations des pertes, trouver la réserve à chaque fois, et établir une distribution empirique des simulations. La situation se complexifie lorsqu'on souhaite modéliser en plus l'erreur d'estimation et de spécification. On devra alors utiliser une généralisation de cet algorithme, appelée *simulations de Monte Carlo par chaînes de Markov*¹¹. On utilise généralement l'algorithme de Metropolis-Hastings (Chib et Greenberg, 1995) pour simuler la distribution de la réserve, comme illustré dans (Verrall et England, 2005) et dans (Wüthrich et Merz, 2008). Cet algorithme est computationnellement plus lourd que les autres. En revanche, il offre une très grande flexibilité lors de l'élaboration du modèle et des hypothèses. La base statistique plus élaborée d'un tel algorithme permet aussi de tester la validité des paramètres.

11. Markov Chain Monte Carlo Simulation (MCMC)

3.2 Analyse des résultats

Plusieurs modèles stochastiques ont été appliqués sur les données de l'industrie automobile Suisse. Les modèles Poisson surdispersé, gamma et Tweedie sont d'abord introduits successivement, et sont ensuite comparés dans la section 3.2.4. Notons que le modèle de Poisson sans surdispersion n'est pas considéré car le contexte des réserves actuarielles s'y prête mal. D'autres détails sur la nature du risque sont disponibles dans le chapitre 4. Par ailleurs, il est préférable de commencer directement avec le modèle surdispersé, et si le paramètre de surdispersion estimé semble près de 1, alors on pourra considérer plus sérieusement le modèle sans surdispersion.

3.2.1 Modèle de Poisson surdispersé

En reprenant la procédure et la définition d'une loi de Poisson surdispersée telle que proposé dans la section 3.1.9, nous avons alors :

$$\begin{aligned} C_{i,j} &\sim \text{Poisson}^*(\lambda_{i,j}, \phi), \lambda_{i,j} = e^{X_{i,j}\beta}, \\ E[C_{i,j}] &= \lambda_{i,j}, \text{Var}[C_{i,j}] = \phi\lambda_{i,j}, \end{aligned}$$

où la moyenne est $\lambda_{i,j}$ et le facteur de surdispersion est ϕ et Poisson^* est une loi de Poisson surdispersée. Les paramètres $\tilde{\mu}$ estimés avec une paramétrisation naturelle sont disponibles dans le tableau 3.1. Le paramètre de surdispersion $\phi = 36722$ est estimé en utilisant le principe de la déviance, après que le vecteur $\tilde{\mu}$ ait été estimé.

Pour trouver les pertes incrémentales, par exemple pour la cellule de la 2^e ligne et de la 11^e colonne (2, 11), on fait le calcul suivant :

$$E[\hat{C}_{2,11}] = \exp\{\mu + \alpha_2 + \beta_{11}\} = \exp\{16.7684 + 0.0257 - 10.9969\} = 329.34$$

Les réserves espérées ainsi que l'EQMP selon la formule analytique sont disponibles dans la section 3.2.4. On peut voir que les pertes espérées entre le modèle de Pois-

Paramètre	Estimé	Écart-type	Prob(> T)
μ	16,7684	0,0380	0,0001
α_2	0,0257	0,0521	0,6247
α_3	0,0410	0,0519	0,4343
α_4	-0,0090	0,0526	0,8643
α_5	-0,0452	0,0531	0,3990
α_6	-0,0444	0,0531	0,4081
α_7	-0,0994	0,0539	0,0721
α_8	-0,1617	0,0550	0,0052
α_9	-0,1602	0,0551	0,0057
β_2	-1,1156	0,0302	0,0001
β_3	-3,2199	0,0765	0,0001
β_4	-4,2218	0,1314	0,0001
β_5	-4,5557	0,1643	0,0001
β_6	-5,4914	0,2806	0,0001
β_7	-5,8743	0,3706	0,0001
β_8	-5,9234	0,4222	0,0001
β_9	-6,8360	0,7658	0,0001
β_{10}	-6,8416	0,9450	0,0001
β_{11}	-10,9969	10,7370	0,3113

Tableau 3.1 Paramètres du modèle de Poisson surdispersé

son surdispersé (tableau 5.5, p.132) et Chain Ladder (tableau 2.1, p. 32) concordent. Les différences sont dues aux erreurs d'arrondissement.

Par ailleurs, on voit que certains paramètres peuvent être considérés comme non-significatifs. Cependant, pour l'instant, très peu de travaux se sont réellement penchés sur cette question et donc, par convention (England et Verrall, 2002), on accepte tous les paramètres, car, intuitivement, on sait qu'il existe un effet d'accident. N'empêche, on peut toujours considérer un modèle plus simple avec moins de paramètres.

3.2.2 Modèle gamma

La loi gamma a été introduite dans la section 1.7.4. Le modèle gamma est très semblable au modèle de Poisson surdispersé. Plus précisément, ce dernier suppose que la variance est proportionnelle à l'espérance, alors que le modèle gamma suppose qu'elle est proportionnelle au carré de l'espérance. Pour le reste, les calculs se font de la même façon.

$$\begin{aligned} C_{i,j} &\sim \Gamma\left(\frac{1}{\phi}; \frac{1}{\phi\lambda_{i,j}}\right), \quad \lambda_{i,j} = e^{X_{i,j}\beta}, \\ E[C_{i,j}] &= \lambda_{i,j}, \quad \text{Var}[C_{i,j}] = \phi\lambda_{i,j}^2. \end{aligned}$$

Les paramètres optimaux sont illustrés dans le tableau 3.2. En utilisant le principe de la déviance pour calculer le paramètre de surdispersion, nous obtenons : $\phi = 0.27678$. Le paramètre de surdispersion est plus petit que celui de la loi de Poisson surdispersée, mais la variance issue du carré de l'espérance contre-balance la statistique. Ce débalancement occasionne une différence majeure entre les deux modèles lorsqu'on s'intéresse à la modélisation de la variance. Une étude plus approfondie de ce phénomène est disponible dans le chapitre 4.

L'encouru cumulatif est disponible dans le tableau 5.7 (p. 134). Puisque la loi

Paramètre	Estimé	Écart-type	Prob(> T)
μ	16,5767	0,2043	0,0001
α_2	0,3279	0,2122	0,1294
α_3	0,1486	0,2107	0,4843
α_4	0,1797	0,2154	0,4086
α_5	0,4913	0,2281	0,0367
α_6	0,0692	0,2348	0,7697
α_7	0,2214	0,2510	0,3825
α_8	0,0005	0,2702	0,9984
α_9	-0,0660	0,3039	0,8290
β_2	-1,1146	0,2104	0,0001
β_3	-3,2275	0,2109	0,0001
β_4	-4,2174	0,2199	0,0001
β_5	-4,6041	0,2326	0,0001
β_6	-5,5045	0,2422	0,0001
β_7	-5,9918	0,2612	0,0001
β_8	-5,9467	0,2837	0,0001
β_9	-6,8834	0,3182	0,0001
β_{10}	-6,8283	0,3651	0,0001
β_{11}	-10,8052	0,4907	0,0001

Tableau 3.2 Paramètres du modèle gamma

gamma est un cas particulier de la classe de lois de Tweedie, on peut se référer aux formules généralisées de la section 3.2.3. Par exemple, pour calculer l'EQMP, on peut utiliser l'équation (3.23) en fixant le paramètre p (expliqué plus loin) à 2.

3.2.3 Modèle de Tweedie

Ce modèle utilise la loi de Tweedie pour modéliser les pertes incrémentales. Cette loi a été nommée d'après M.C.K. Tweedie, qui a introduit cette classe de modèles

dans son article (Tweedie, 1984). Déjà, plusieurs auteurs avaient commencé à analyser cette classe chacun de leur côté : (Morris, 1981), (Bar-Lev et Enis, 1984), (Hougaard, 1986). Aujourd'hui, une des sources les plus complètes dans ce domaine est le livre de (Jørgensen, 1997).

Il a été démontré dans les sections précédentes que le modèle de Poisson surdispersé et le modèle gamma peuvent être exprimés uniquement selon une relation entre l'espérance et la variance. Le modèle de Tweedie est une généralisation de cette relation. En fait, il suppose que la variance est proportionnelle à l'espérance à la puissance p comme dans l'équation suivante :

$$\text{Var}[C_{i,j}] = \phi E[C_{i,j}]^p .$$

Construction du modèle

Dans ce mémoire, on s'intéresse particulièrement à $p \in (1, 2)$ qui est en quelque sorte une interpolation entre le modèle de Poisson (surdispersé) ($p = 1$) et le modèle gamma ($p = 2$). Pour arriver à ces résultats, nous suivons de près la démarche de (Wüthrich, 2003). Son article est l'un des premiers à appliquer ce modèle aux réserves actuarielles. Pour la cellule (i, j) , on utilise les hypothèses initiales suivantes :

- $R_{i,j}$ est le nombre de réclamations ,
- $C_{i,j}^{(k)}$ est la k^e perte incrémentale ,
- $C_{i,j}$ est la perte incrémentale totale ,
- $C_{i,j} = \sum_{k=1}^{R_{i,j}} C_{i,j}^{(k)}$ où $R_{i,j}$ et $C_{i,j}^{(k)}$ sont tous indépendants entre eux ,
- $R_{i,j}$ suit une loi de Poisson de moyenne $w_{i,j}\vartheta_{i,j}$,
- $C_{i,j}^{(k)}$ suit une loi gamma de moyenne $\tau_{i,j}$ et variance $\frac{\tau_{i,j}^2}{\nu}$,
- $Y_{i,j} = C_{i,j}/w_{i,j}$ où $w_{i,j}$ est l'exposition .

En d'autres mots, nous avons :

$$C_{i,j} = C_{i,j}^{(1)} + C_{i,j}^{(2)} + \dots + C_{i,j}^{(R_{i,j})} .$$

(Wüthrich, 2003) s'intéresse à trouver la densité de Y . Pour ce faire, on trouve la densité jointe du modèle de somme composée des variables aléatoires C et R (les indices sont sous-entendus) :

$$\begin{aligned}
 f_{R,Y}(r, y|\vartheta, \tau, v)dc &= f_{R,C}(r, wy|\vartheta, \tau, v)dc \\
 &= f_C(wy|R=r)P(R=r)wdy \\
 &= \frac{(\nu/\tau)^{\nu r}}{\Gamma(\nu r)}(wy)^{\nu r-1} \exp\left\{-\frac{\nu}{\tau}wy\right\} \frac{(w\vartheta)^r}{r!} \exp\{-w\vartheta\}wdy,
 \end{aligned} \tag{3.10}$$

où $\frac{(\nu/\tau)^{\nu r}}{\Gamma(\nu r)}(wy)^{\nu r-1} \exp\left\{-\frac{\nu}{\tau}wy\right\}$ est la densité d'une loi gamma qui représente la somme de r lois gammas indépendantes, et $\exp\{-w\vartheta\} \frac{(w\vartheta)^r}{r!}$ est la fonction de probabilité de la loi de Poisson évaluée à r . L'équation (3.10) devient alors :

$$f_{R,Y}(r, y|\vartheta, \tau, v)dc = \left(\vartheta \left(\frac{\nu}{\tau}\right)^{\nu} y^{\nu} w^{\nu+1}\right)^r \frac{1}{r!\Gamma(\nu r)y} \exp\left\{\frac{w}{\phi} \left(-\frac{\nu\phi}{\tau}y - \vartheta\phi\right)\right\} dy. \tag{3.11}$$

Il s'avère qu'en reparamétrisant judicieusement, on peut reconnaître l'appartenance du modèle à la *famille exponentielle linéaire*. En effet, si on pose :

$$\begin{aligned}
 p &= \frac{\nu+2}{\nu+1}, \quad p \in (1, 2), \\
 \mu &= \vartheta\tau, \\
 \phi &= \frac{\vartheta^{1-p}\tau^{2-p}}{(2-p)},
 \end{aligned}$$

on obtient alors $\nu = (2-p)/(p-1)$, et finalement l'équation (3.11) devient :

$$f_{R,Y}(r, y|\vartheta, \tau, v)dy = \left(\frac{y^{\nu}(w/\phi)^{\nu+1}}{(p-1)^{\nu}(2-p)}\right)^r \frac{1}{r!\Gamma(\nu r)y} \exp\left\{\frac{w}{\phi} \left(-y \frac{\mu^{1-p}}{1-p} - \frac{\mu^{2-p}}{2-p}\right)\right\} dy \tag{3.12}$$

et peut être ainsi exprimé sous la forme :

$$f_{R,Y}(r, y|\vartheta, \tau, v)dy = c\left(y; \frac{w}{\phi}; p\right) \exp\left\{\frac{w}{\phi}(y\theta - \kappa_p(\theta))\right\} dy,$$

où, lorsque $r > 0$:

$$\begin{aligned}\theta &= \theta(\mu) = \frac{\mu^{1-p}}{1-p} < 0 \\ \kappa_p(\theta) &= \frac{\mu^{2-p}}{2-p} = \frac{1}{2-p} ((1-p)\theta)^{\frac{2-p}{1-p}} \\ c\left(y; \frac{w}{\phi}; p\right) &= \left(\frac{y^\nu(w/\phi)^{\nu+1}}{(p-1)^\nu(2-p)}\right)^r \frac{1}{r!\Gamma(\nu r)y}.\end{aligned}\quad (3.13)$$

et quand $R = 0$, la fonction de densité de Y possède un point de masse défini comme suit :

$$P(R = 0) = \exp\{-w\vartheta\} = \exp\left\{\frac{w}{\phi}(-\kappa_p(\theta))\right\}.$$

La loi inconditionnelle de Y doit tenir compte de tous les cas possibles du nombre de réclamations. Ainsi, on doit sommer sur toutes les réalisations possibles de R . On obtient alors :

$$f_Y(y|\vartheta, \tau, v)dy = \sum_r \left(\frac{y^\nu(w/\phi)^{\nu+1}}{(p-1)^\nu(2-p)}\right)^r \frac{1}{r!\Gamma(\nu r)y} \exp\left\{\frac{w}{\phi}\left(-y\frac{\mu^{1-p}}{1-p} - \frac{\mu^{2-p}}{2-p}\right)\right\} dy. \quad (3.14)$$

Cette somme infinie doit être approximée numériquement. D'autres détails sont disponibles dans (Peters, Shevchenko et Wüthrich, 2009). Afin de simplifier la démarche, puisque nous connaissons le nombre de réclamations, nous allons substituer la valeur connue r à la variable aléatoire R (la lettre r était utilisée comme variable de comptage précédemment sera désormais remplacée comme variable représentant une fréquence observée) et donc l'équation (3.14) se simplifie en :

$$f_Y(y|\vartheta, \tau, v)dy = \left(\frac{y^\nu(w/\phi)^{\nu+1}}{(p-1)^\nu(2-p)}\right)^r \frac{1}{r!\Gamma(\nu r)y} \exp\left\{\frac{w}{\phi}\left(-y\frac{\mu^{1-p}}{1-p} - \frac{\mu^{2-p}}{2-p}\right)\right\} dy.$$

Par ailleurs, en utilisant l'équation (3.12) On peut identifier que le modèle fait partie de la famille exponentielle linéaire, avec moyenne $\mu_{i,j}$, fonction de variance avec puissance p et paramètre de surdispersion $\frac{w_{i,j}}{\phi}$, comme illustré ici :

$$Y_{i,j} \sim \text{ED}\left(\mu_{i,j} = \frac{\lambda_{i,j}}{w_{i,j}}; p; \frac{\phi}{w_{i,j}}\right),$$

où ED signifie la famille exponentielle linéaire. L'ouvrage de (Jørgensen, 1997) démontre que cette classe de modèles établit une relation entre l'espérance et la variance de la loi qui est illustrée dans l'équation (3.16). De plus, ce même ouvrage stipule que la moyenne et la variance du modèle peuvent se calculer en utilisant les équations (3.15) et (3.16) illustrées ici :

$$E[Y_{i,j}] = \kappa'_p(\theta_{i,j}) = \frac{\partial(\kappa_p(\theta_{i,j}))}{\partial\theta_{i,j}} = \mu_{i,j} , \quad (3.15)$$

$$\text{Var}[Y_{i,j}] = \frac{\phi}{w_{i,j}} \kappa''_p(\theta_{i,j}) = \frac{\phi}{w_{i,j}} \mu_{i,j}^p . \quad (3.16)$$

Ces relations permettent à la fois de calculer aisément les deux premiers moments de la loi pour pouvoir utiliser l'algorithme de Newton-Raphson, et aussi d'établir une procédure générale qui s'applique à de nombreux modèles qui ont une relation entre l'espérance et la variance, tel que le modèle normal, Poisson surdispersé ou gamma.

Fonction de vraisemblance

En utilisant la densité de l'équation (3.11), la fonction de log-vraisemblance est alors :

$$l = \sum_{i,j} \left(r_{i,j} \log \left(\frac{(w_{i,j}/\phi)^{\nu+1} y_{i,j}^\nu}{(p-1)^\nu (2-p)} \right) - \log(r_{i,j}! \Gamma(r_{i,j}\nu) y_{i,j}) + \frac{w_{i,j}}{\phi} \left(y_{i,j} \frac{\mu_{i,j}^{1-p}}{1-p} - \frac{\mu_{i,j}^{2-p}}{2-p} \right) \right) . \quad (3.17)$$

Rappelons que la fonction de vraisemblance (3.17) est conditionnelle au nombre de réclamations $r_{i,j}$. Cette situation nous permet d'estimer à la fois les paramètres $\tilde{\mu}$ associés aux moyennes, le paramètre de surdispersion ϕ et le paramètre p .

À l'inverse, lorsque le nombre de réclamations est manquant, il on ne peut estimer p qu'en utilisant une série d'approximations comme démontré dans (Peters, Shevchenko et Wüthrich, 2009). Cette situation n'est pas considérée dans ce mémoire. Par ailleurs, toujours lorsque le nombre de réclamations est manquant, si on ne souhaite pas utiliser

ces approximations, les moyennes peuvent toujours être estimées, mais au prix de l'hypothèse supplémentaire que p est connu et fixé. Ainsi, sachant p fixé, et n'ayant que μ comme paramètre d'intérêt, la formule (3.17) devient :

$$l = \sum_{i,j} \frac{w_{i,j}}{\phi} \left(y_{i,j} \frac{\mu_{i,j}^{1-p}}{1-p} - \frac{\mu_{i,j}^{2-p}}{2-p} \right), \quad (3.18)$$

où tous les autres termes de l'équation (3.17) sont ignorés car ils ne dépendent pas de $\mu_{i,j}$.

Relations entre Y et C

La densité de C étant une somme composée Poisson-gamma, on en déduit que C est aussi une loi de Tweedie. De plus, on peut trouver ses premiers moments de la façon suivante :

$$\begin{aligned} Y_{i,j} &\sim \text{Tweedie} \left(\mu_{i,j} ; p ; \frac{\phi}{w_{i,j}} \right), \\ E[Y_{i,j}] &= \mu_{i,j}, \\ \text{Var}[Y_{i,j}] &= \frac{\phi}{w_{i,j}} \mu_{i,j}^p, \end{aligned}$$

où $\mu_{i,j}$ est la moyenne, $\frac{\phi}{w_{i,j}}$ est le paramètre de surdispersion et p est la puissance de la relation entre l'espérance et la variance. On en déduit que :

$$\begin{aligned} E[C_{i,j}] &= E[w_{i,j} Y_{i,j}] = \lambda_{i,j}, \\ \text{Var}[C_{i,j}] &= \text{Var}[w_{i,j} Y_{i,j}] = w_{i,j}^2 \frac{\phi}{w_{i,j}} \mu_{i,j}^p = \phi w_{i,j} \mu_{i,j}^p. \end{aligned} \quad (3.19)$$

Signification du paramètre p

La démonstration présentée ici ne attarde qu'au cas où p est entre 1 et 2. (Jørgensen, 1997) indique que la classe des modèles de Tweedie s'étend sur un grand intervalle de valeurs réelles. Le tableau 3.3 (p. 79) résume les modèles pour toutes les valeurs de p .

On y voit que plusieurs lois présentées jusqu'ici sont des cas particuliers de la loi de Tweedie. Donc, toutes les formules du modèle général seront applicables aux cas particuliers aussi.

Loi	p	Domaine de C	Domaine de λ
Stable extreme	$p < 0$	\mathbf{R}	\mathbf{R}_+
Normale	$p = 0$	\mathbf{R}	\mathbf{R}
[N'existe pas]	$0 < p < 1$	-	\mathbf{R}_+
Poisson	$p = 1$	\mathbf{N}_0	\mathbf{R}_+
Somme Poisson composée	$1 < p < 2$	\mathbf{R}_0	\mathbf{R}_+
Gamma	$p = 2$	\mathbf{R}_+	\mathbf{R}_+
Stable positive	$2 < p < 3$	\mathbf{R}_+	\mathbf{R}_+
Inverse gaussienne	$p = 3$	\mathbf{R}_+	\mathbf{R}_+
Stable positive	$p > 3$	\mathbf{R}_+	\mathbf{R}_+
Stable extreme	$p \rightarrow \infty$	\mathbf{R}	\mathbf{R}

Tableau 3.3 Sommaire des modèles exponentiels de la classe de Tweedie

Paramètre de surdispersion

On peut estimer le paramètre de surdispersion par au moins deux façons. Tout d'abord, pour utiliser le principe du maximum de vraisemblance, on estime le facteur de surdispersion ϕ en trouvant la solution de la racine de la première dérivée de la fonction de vraisemblance de l'équation (3.17). On obtient alors :

$$\widehat{\phi_p} = \frac{-\sum_{i,j} w_{i,j} \left(y_{i,j} \frac{\mu_{i,j}^{1-p}}{1-p} - \frac{\mu_{i,j}^{2-p}}{2-p} \right)}{(1+\nu) \sum_{i,j} r_{i,j}} . \quad (3.20)$$

où ϕ_p indique que ϕ dépend de p et doit être conjointement optimisé. Le paramètre de surdispersion peut être optimisé conjointement avec les paramètres $\tilde{\mu}$ associés aux moyennes, ou bien en alternance, comme proposé dans (Wüthrich, 2003).

Par ailleurs, le paramètre de surdispersion calculé selon le principe de la déviance tel que défini dans la section 3.1.9 est :

$$\widehat{\phi}_p = \sum_{i,j} \frac{2}{N-Q} \left(y_{i,j} \frac{y_{i,j}^{1-p} - \mu_{i,j}^{1-p}}{1-p} - \frac{y_{i,j}^{2-p} - \mu_{i,j}^{2-p}}{2-p} \right), \quad (3.21)$$

où N est le nombre d'observations et Q le nombre de paramètres dans le modèle. Notons que dans ce cas, contrairement à l'estimation par maximum de vraisemblance de ϕ , il n'est pas nécessaire de connaître le nombre de réclamations.

Optimisation du paramètre p

Tel qu'indiqué dans les équations (3.20) et (3.21), le facteur de surdispersion dépend du paramètre p , et ces deux derniers ont aussi une influence sur les moyennes μ . Il faut donc construire un algorithme qui optimise tous ces paramètres à la fois. Par ailleurs, notons que dans tous les cas considérés dans ce mémoire, lorsqu'on optimise p , on suppose connaître le nombre de réclamations.

L'article de (Wüthrich, 2003) met en évidence une orthogonalité au niveau de l'optimisation entre les moyennes μ et ϕ ainsi qu'une autre orthogonalité entre μ et p . En utilisant ces résultats, on peut se permettre d'optimiser en alternance les moyennes et les deux autres paramètres, obtenant une convergence après quelques itérations seulement. En quelques mots, l'algorithme se résume ainsi :

1. Fixer p à une valeur connue, par exemple $p = 1.5$,
2. En gardant p fixé, on optimise les paramètres correspondant aux moyennes à l'aide d'un modèle linéaire généralisé,
3. En supposant que les moyennes sont fixées, on optimise en même temps p et ϕ_p ,
4. On alterne les étapes 2 et 3 jusqu'à convergence.

Itération k	0	1	2	3	4
$p^{(k)}$	1.5000	1.1742	1.1740	1.1740	1.1740
Réserves espérées	1 424 587	1 440 103	1 440 113	1 440 113	1 440 113

Tableau 3.4 Estimation de p avec le principe de vraisemblance maximale

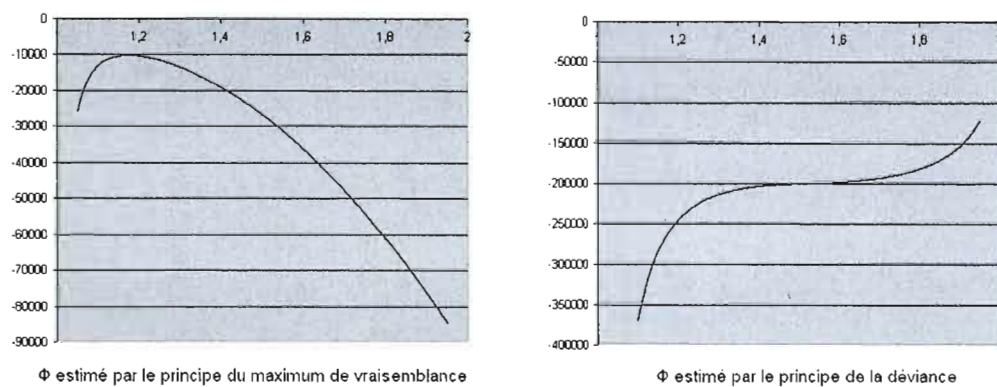


Figure 3.3 Fonction de vraisemblance de p à la première itération

Le tableau 3.4 montre la trace de l'algorithme pour le principe de vraisemblance maximale, tel qu'utilisé dans (Wüthrich, 2003). On voit que la convergence se fait très rapidement.

La figure 3.3 illustre la fonction de log-vraisemblance (3.17), au début de la deuxième itération de l'algorithme, en fonction de p , lorsque ϕ est calculé selon le principe de vraisemblance ou le principe de déviance. Il s'avère que dans cet exemple-ci, le paramètre p ne peut être optimisé lorsqu'on utilise le principe de la déviance. Pour l'instant, les raisons de cette divergence restent inconnues.

Il est intéressant de mentionner qu'un logiciel de statistique tel que SAS peut optimiser directement tous les paramètres à la fois en utilisant implicitement le principe de

vraisemblance maximale pour évaluer ϕ . Ainsi, la programmation s'en retrouve quelque peu allégée.

Erreur quadratique moyenne de la réserve

En ajoutant l'erreur d'estimation, telle que définie à la section 3.1.11, à l'erreur statistique, telle que calculée dans l'équation (3.19), on obtient :

$$\begin{aligned} \text{Erreur totale} &= \text{Erreur statistique} + \text{Erreur d'estimation} , \\ \text{EQMP}[\hat{C}_{i,j}] &= \phi w_{i,j} \mu_{i,j}^p + (w_{i,j} \mu_{i,j})^2 \text{Var}[\eta_{i,j}] . \end{aligned}$$

Finalement, pour calculer l'erreur de la réserve totale,

$$\begin{aligned} \text{EQMP}[R] &= \sum_{(i,j) \in \Delta} \phi w_{i,j} \mu_{i,j}^p \\ &+ \sum_{(i,j) \in \Delta} (w_{i,j} \mu_{i,j})^2 \text{Var}[\eta_{i,j}] \\ &+ \sum_{(i,j) \in \Delta, (i_1,j_1) \neq (i_2,j_2)} (w_{i_1,j_1} \mu_{i_1,j_1})(w_{i_2,j_2} \mu_{i_2,j_2}) \text{Cov}(\eta_{i_1,j_1}, \eta_{i_2,j_2}) . \end{aligned} \tag{3.22}$$

Il s'avère que l'erreur statistique de l'EQMP dépend de ϕ . Hélas, pour les données analysées ici, avec des poids uniformes et une exposition stable selon l'année d'accident, le facteur de surdispersion selon le principe de la déviance est de l'ordre de 30000, alors que celui de la vraisemblance maximale est de l'ordre de 1500. Cette différence marquée entraîne un impact significatif sur l'incertitude de la réserve.

Cette inconsistance n'est pas documentée dans la littérature scientifique. Par convention, la plupart des articles, comme par exemple (Wüthrich, 2003), utilisent le facteur de surdispersion issu du principe de la déviance pour calculer l'EQMP alors qu'ils utilisent le principe du maximum de vraisemblance pour estimer p . Cette approche soulève néanmoins quelques réticences et le sujet mérite d'être analysé plus en détails dans les recherches futures.

Résultats du modèle de Tweedie

Le tableau 3.5 montre les paramètres correspondant aux moyennes pour un modèle avec w qui est le volume d'affaire. L'ordre de grandeur du paramètre de base μ représentant la cellule (1,1) est différent de ceux des modèles Poisson surdispersé et gamma car ici on modélise les $Y_{i,j}$ au lieu des $C_{i,j}$. Quant aux autres paramètres, leur comportement est en général très semblable à ceux des autres modèles.

Le tableau 5.8 (p. 135) indique l'encours cumulatif du modèle de Tweedie selon l'année d'accident et la période de développement. Ces résultats se rapprochent de ceux des autres modèles.

3.2.4 Comparaison des modèles stochastiques

Le tableau 3.6 montre les réserves et les EQMP des modèles stochastiques. Le modèle de Poisson et le modèle gamma ne tiennent pas compte de l'exposition, alors que le modèle de Tweedie pondère par le volume d'affaire. On utilise le principe de vraisemblance maximale pour optimiser p . L'EQMP est calculée avec un paramètre de surdispersion calculé selon le principe de la déviance.

Les réserves du modèle de Tweedie sont quelque part entre celles du modèle de Poisson et du modèle gamma. Les trois modèles ont comme résultat des réserves semblables pour toutes les années. Quant à l'erreur quadratique moyenne, les résultats du modèle gamma sont inférieurs à ceux des autres modèles. Le chapitre suivant propose une explication à ce phénomène.

Paramètre	Estimé	Écart-type	Prob(> T)
μ	5,1435	0,0506	0,0001
α_2	0,0373	0,0690	0,5913
α_3	0,1007	0,0689	0,1512
α_4	0,0800	0,0698	0,2581
α_5	0,0862	0,0704	0,2272
α_6	0,0436	0,0707	0,5409
α_7	0,0700	0,0717	0,3342
α_8	0,0256	0,0733	0,7283
α_9	0,0539	0,0739	0,4699
β_2	-1,1153	0,0394	0,0001
β_3	-3,2200	0,0821	0,0001
β_4	-4,2223	0,1284	0,0001
β_5	-4,5580	0,1557	0,0001
β_6	-5,4936	0,2445	0,0001
β_7	-5,8798	0,3122	0,0001
β_8	-5,9238	0,3534	0,0001
β_9	-6,8404	0,5918	0,0001
β_{10}	-6,8463	0,7277	0,0001
β_{11}	-11,0067	5,7397	0,0617

Tableau 3.5 Paramètres du modèle Tweedie, avec $\phi = 29348$ (déviante)

Année	w_i	Poisson		Tweedie (p=1,1741)		gamma	
		R_i	EQMP ^{1/2}	R_i	EQMP ^{1/2}	R_i	EQMP ^{1/2}
1	112 953	-	-	-	-	-	-
2	110 364	329	4 950	326	2 638	446	350
3	105 400	21 663	34 813	21 565	26 804	20 244	13 548
4	102 067	41 007	46 119	40 716	35 556	40 282	19 874
5	99 124	88 557	65 305	89 298	53 297	122 600	54 934
6	101 460	140 148	80 882	138 335	66 052	122 745	49 165
7	94 753	204 154	95 858	204 262	80 906	223 256	86 786
8	92 326	363 095	125 632	360 484	111 999	337 425	140 283
9	89 545	603 156	161 248	597 056	150 003	533 920	232 187
Total	-	1 462 108	317 610	1 452 042	271 886	1 400 887	318 483
ϕ		36721		29348		0.27878	

Tableau 3.6 Sommaire des différents modèles (l'exposition est appliquée seulement au modèle de Tweedie)

CHAPITRE IV

MODÉLISATION DE LA VARIANCE

4.1 Problématique

Les modèles présentés dans ce chapitre ont pour but de s'adapter à des situations où le modèle Chain Ladder obtient des résultats inappropriés. Les insuffisances de ce dernier ont été présentées dans la section 2.1. En général, la situation typique veut que la large part des réclamations se fasse dans les périodes jeunes, et que la sévérité moyenne, ainsi que sa variance, augmente avec la période de développement. Ce débalancement suscite beaucoup de difficultés à capter correctement la nature du risque si on modélise uniquement la moyenne des coûts totaux. Par exemple, en assurance-automobile pour les blessures corporelles¹, le temps avant que le sinistre soit réglé lorsqu'il y a appel en justice peut varier. La plupart des cas se règlent rapidement. Cependant, il reste des cas difficiles qui prennent plus de temps. Généralement, le conflit perdure lorsque le montant en jeu est plus important. Ainsi, il peut arriver que ces cas se concluent par aucune indemnisation ou bien une facture très salée pour l'assureur.

Cette observation a déjà été faite par (Adler et Kline Jr, 1978). Son modèle s'aligne avec l'idée de modéliser séparément la fréquence et la sévérité. C'est une avenue en plein développement qui recèle encore beaucoup de potentiel inexploité. Cependant, les modèles présentés ici se basent sur une approche qui analyse les coûts totaux seule-

1. Automobile Bodily Injury (AB) & Third Party Liabilities (TPL)

ment.

On appelle *coûts totaux* la somme des réclamations pour une année et une période fixée. La *fréquence* désigne le nombre de réclamations, alors que la *sévérité* représente le montant de perte par réclamation. La *prime pure* représente les coûts totaux moyens, c'est-à-dire, la fréquence moyenne multipliée par la sévérité moyenne. Afin de bien comprendre la nature du risque, comme pour le modèle de Tweedie avec ($1 < p < 2$), nous supposons qu'un modèle *somme composée* reflète adéquatement les relations entre la fréquence, la sévérité et les coûts totaux, comme indiqué ici :

$$S = \sum_{k=1}^N X_k ,$$

où N modélise la fréquence et X_k , la sévérité, et N, X_1, X_2, \dots, X_N sont indépendants entre eux, avec la convention que $S = 0$ si $N = 0$. Tous les X_k suivent la même loi qui possède une espérance et variance finie. Nous avons alors :

$$E[S] = E[N]E[X] , \text{ Var}[S] = E[N]\text{Var}[X] + \text{Var}[N]E[X]^2 .$$

Les modèles de ce chapitre ne modélisent que le triangle des coûts totaux, alors qu'en réalité, il y a deux tendances sous-jacentes, une pour la fréquence, et une pour la sévérité. Pour faire face à cette situation, on peut ajouter une composante de flexibilité supplémentaire au modèle en modélisant la variance en même temps que les moyennes. Ainsi, le modèle possède assez de flexibilité pour capter implicitement à la fois la tendance de la fréquence et celle de la sévérité. Ainsi, lorsque la variance est élevée, on peut déterminer si celle-ci est due à un coût total moyen élevé, ou bien à une sévérité moyenne élevée. Considérons l'exemple du tableau 4.1. Nous supposons que la fréquence suit une loi de Poisson, et la sévérité, une loi exponentielle. Le premier cas représente une situation où on a en moyenne 10 réclamations et une sévérité moyenne de 10. Si on double la fréquence moyenne, comme démontré dans le deuxième cas, alors la prime pure est aussi doublée et la variance des coûts totaux aussi. En revanche, si on double la sévérité moyenne, la variance des coûts totaux quadruple, alors que la prime pure ne

	Cas 1	Cas 2	Cas 3
$E[N]$	10	20	10
$Var[N]$	10	20	10
$E[X]$	10	10	20
$Var[X]$	100	100	400
$E[S]$	100	200	200
$Var[S]$	2000	4000	8000

Tableau 4.1 La moyenne et la variance des coûts totaux selon différentes situations

fait que doubler, comme illustré dans le troisième cas.

Un modèle qui suppose la même variance ne pourra modéliser correctement les cas 2 et 3. Les modèles de cette section ajoutent une variable supplémentaire qui multiplie les variances des coûts, permettant de distinguer les cas 2 et 3 grâce à l'information provenant de la sévérité et de la fréquence. En pratique, un triangle typique possède une fréquence élevée au début avec une sévérité petite, et une petite fréquence et grande sévérité aux périodes avancées. Ainsi, il faut ajuster à la hausse la variance de la prime pure des périodes avancées.

Par ailleurs, le choix du modèle a un impact important sur la structure de la variance. En effet, dans une régression classique, certains modèles accordent naturellement une variance des coûts totaux plus élevée à mesure que la prime pure augmente, comme expliqué dans la section 3.2.3. Par exemple, la loi de Poisson possède une variance proportionnelle à la moyenne, alors que pour la loi gamma, elle l'est au carré de la moyenne. La loi normale est un cas où la variance n'est pas affectée naturellement par la moyenne.

Supposons qu'on régresse une courbe sur trois observations selon les modèles normal, Poisson et gamma. Notons que les moyennes sont reliées à la variance à un

$E[S]$	1	2	3
Normal	ϕ_1	ϕ_1	ϕ_1
Poisson	ϕ_2	$2\phi_2$	$3\phi_2$
Gamma	ϕ_3	$4\phi_3$	$9\phi_3$

Tableau 4.2 La variance ($\text{Var}[S]$) selon différentes moyennes ($E[S]$), pour les modèles normal, Poisson et gamma

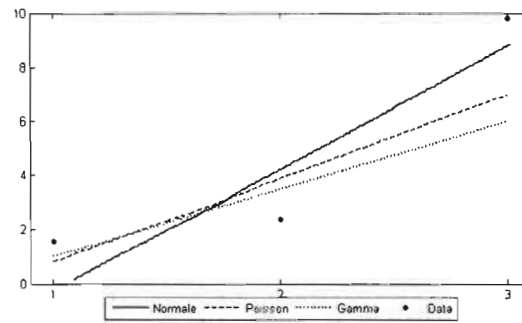


Figure 4.1 Courbe optimale de régression selon le modèle normal, Poisson ou gamma

facteur près, qui est aussi appelé paramètre de surdispersion (section 3.1.9). Chaque loi possède son propre paramètre de surdispersion, noté ϕ_{loi} , et il doit être estimé au même titre que les paramètres $\tilde{\mu}$ relatifs aux moyennes. Supposons qu'après optimisation, nous obtenons les valeurs du tableau 4.2.

Le nombre qui multiplie le facteur de surdispersion est ce qu'on appelle ici la *variance naturelle*. On remarque que pour la loi gamma, la variance naturelle prend des proportions énormes à mesure que $E[S]$ augmente. On conclut alors que la loi gamma accorde naturellement plus de variance aux données avec moyenne élevée, comparative-ment aux autres lois. Le graphique 4.1 tiré de (Anderson et al., 2004) illustre bien la situation.

Une variabilité moindre implique que la courbe est plus susceptible de passer près

du point d'observation. Ce principe explique pourquoi la courbe normale essaie de passer aussi près de toutes les observations, alors que la loi de Poisson et la loi gamma essaient de se rapprocher davantage des données avec moyenne petite.

En pratique, il serait intéressant d'accorder plus d'importance aux données jeunes, et moins d'importance aux données provenant des périodes avancées, puisque le rapport de la variance des coûts totaux par rapport à son espérance risque d'être plus élevé pour celles-ci. En d'autres mots, le modèle doit être en mesure de pouvoir gonfler la variance pour les données matures, et inversement réduire la variance pour les données des périodes du début. Cette affirmation soulève un paradoxe sur l'utilisation de la loi gamma. On dit souvent que la loi gamma accorde une variabilité naturelle plus élevée aux données. Donc, on pourrait penser à l'utiliser pour donner plus d'importance à la probabilité que la réserve finale soit extrême. Contre toute attente, c'est exactement l'effet inverse qui se produit. Puisque la prime pure est faible pour les périodes avancées, en raison de la fréquence réduite, alors la loi gamma accorde beaucoup plus d'importance à ces données, donc la courbe optimale de régression passera plus proche. À l'inverse, pour les données des périodes jeunes, puisque la prime pure est élevée en raison de la fréquence élevée, alors la variabilité naturelle sera énorme, et donc la courbe optimale de régression peut passer loin des données observées.

On constate donc que la loi de Poisson peut sous-estimer les pertes pouvant survenir aux périodes avancées car elle accorde beaucoup d'importance aux données observées dans ces périodes. Et comme ce modèle est intrinsèquement relié à la méthode Chain Ladder, on comprend que cette insuffisance rend nécessaire l'utilisation de modèles à ailes relevées. Suivant la même logique, on constate que l'effet de la loi gamma est encore plus prononcé. Ainsi, on en conclut que le choix de la loi a un impact significatif sur la modélisation de la variance.

Comme le sujet est encore en plein développement, il n'y a pour l'instant aucun article scientifique qui applique directement les méthodes présentées dans ce chapitre dans un contexte de réserves actuarielles. Les articles référés ici se rapportent soit à des situations théoriques, soit à des applications dans un autre domaine pratique.

Définition du modèle pour les coûts totaux

Le modèle utilise les hypothèses et les notations du chapitre 3. L'innovation consiste à introduire un terme plus général au niveau de la variance. Par exemple, pour un modèle de Tweedie, nous avons :

$$E[C_{i,j}] = \lambda_{i,j} , \text{ Var}[C_{i,j}] = \phi \lambda_{i,j}^p .$$

On veut admettre une structure de variance plus flexible. Alors, le nouveau modèle est défini comme suit :

$$E[C_{i,j}] = \lambda_{i,j} , \text{ Var}[C_{i,j}] = \phi_{i,j} \lambda_{i,j}^p , \quad (4.1)$$

où $\phi_{i,j}$ est le paramètre de surdispersion associé à la cellule (i,j) .

Rappelons que dans la section 4.1, nous avons vu que les tendances par année de développement entre la fréquence et la sévérité sont différentes, ce qui n'est pas nécessairement le cas par année d'accident. C'est pourquoi, dans les exemples qui suivent, on suppose que seules les colonnes ont un effet sur la surdispersion, donc : $\phi_{i,j} = \phi_j, \forall i, \forall j$. Cependant, nous allons néanmoins conserver la notation $\phi_{i,j}$ pour rappeler que les méthodes présentées peuvent être appliquées à la fois aux lignes et aux colonnes. Par ailleurs, en ce qui concerne les moyennes, on maintient l'idée que les années d'accident ont un effet tout comme les périodes de développement.

Si on souhaite incorporer une tendance dans le modèle, on peut réécrire l'équation

(4.1) de la façon suivante :

$$E[C_{i,j}] = \lambda_{i,j} , \text{ Var}[C_{i,j}] = \phi_{i,j} \lambda_{i,j}^p , \phi_{i,j} = \phi f(j) ,$$

où ϕ est le paramètre de surdispersion comme dans le modèle traditionnel. La fonction $f(j)$ représente l'ajustement de la variance de la colonne j . Par exemple, si on souhaite avoir une variance deux fois plus grande que pour le modèle classique pour la période 2, on fixe $f(2) = 2$. Ensuite, si on veut tripler la variance de la troisième période de développement, on fixe $f(3) = 3$. La valeur de la fonction $f(j)$ est aussi appelée *rapport de variances*.

On peut utiliser notre jugement pour déterminer quelles sont les valeurs les plus plausibles de la fonction $f(j)$. Néanmoins, on peut aussi estimer les rapports de variance directement à partir des données.

La section 4.2 se décompose en deux parties. On propose d'abord un modèle de dispersion qui estime automatiquement les facteurs de surdispersion de chaque colonne. Ensuite, on introduit un modèle de dispersion qui permet d'incorporer une tendance dans la variance. La section 4.3 arbore une approche différente basée sur les modèles linéaires généralisés doubles. Par la suite, on compare les résultats de ces deux séries de modèles. Finalement, la dernière section est consacrée à l'énumération de divers problèmes pratiques reliés aux modèles présentés ici et quelques pistes de solutions pour les développements futurs.

4.2 Modèles de dispersion

Dans cet ouvrage, on entend par *modèle de dispersion*, un modèle qui utilise le principe de vraisemblance maximale pour estimer un ou plusieurs paramètres reliés au facteur de surdispersion. (Wüthrich, 2003) présente l'idée de modéliser les facteurs de surdispersion en fonction des lignes ou des colonnes, mais ne met pas son idée en application. Il se réfère à l'article de (Smyth et Jorgensen, 2002) qui développe une solution

élégante, qui sera présentée dans la section 4.3. Les idées de cette section ne se basent sur aucun article extérieur, et ont été développées de façon indépendante.

En reprenant les mêmes hypothèses et notations de la section 3.2.3, pour un modèle de Tweedie avec exposition, nous avons :

$$\begin{aligned} Y_{i,j} &\sim \text{Tweedie} \left(\mu_{i,j} ; p ; \frac{\phi_{i,j}}{w_{i,j}} \right) , \\ E[Y_{i,j}] &= \mu_{i,j} , \quad \text{Var}[Y_{i,j}] = \frac{\phi_{i,j}}{w_{i,j}} \mu_{i,j}^p , \end{aligned}$$

et aussi :

$$\begin{aligned} Y_{i,j} &= \frac{C_{i,j}}{w_{i,j}} , \\ E[C_{i,j}] &= \lambda_{i,j} , \quad \text{Var}[C_{i,j}] = \phi_{i,j} w_{i,j} \mu_{i,j}^p . \end{aligned}$$

4.2.1 Estimation des rapports de variance en utilisant le principe de vraisemblance maximale

Comme introduit dans la section 3.1.9, l'idée est de maximiser la fonction de log-vraisemblance en optimisant en même temps les paramètres $\tilde{\mu}$ relatifs aux moyennes et les paramètres relatifs aux rapports de variances. Pour ce faire, définissons V et $\tilde{\gamma}$ tel que :

$$\phi_{i,j} = \exp\{V_{i,j}\tilde{\gamma}\} ,$$

où la matrice V est la matrice catégorielle des paramètres associés aux variances, et le vecteur $\tilde{\gamma} = \{\phi, \phi_2, \dots, \phi_J\}$ indique les paramètres des colonnes associés à la modélisation de ϕ . Le vecteur $\tilde{\gamma}$ peut aussi contenir des paramètres relatifs aux années d'accident sans perte de généralité, mais dans ce mémoire, nous ne considérons que l'effet des périodes de développement. Ainsi, pour un modèle de Tweedie avec $p \in (1, 2)$, l'équation (3.17), qui représente la fonction de log-vraisemblance, se généralise en :

Période	1	2	3	4	5	6	7	8	9	10	11
ϕ_j	240	1467	408	2 396	6 725	15 449	25 497	50 346	66 313	84 833	84 833

Tableau 4.3 Paramètres de surdispersion selon la période, $p = 1.8111$

$$l = \sum_{i,j} r_{i,j} \log \left(\frac{(w_{i,j}/\phi_{i,j})^{\nu+1} y_{i,j}^{\nu}}{(p-1)^{\nu}(2-p)} \right) - \log(r_{i,j}! \Gamma(r_{i,j}\nu) y_{i,j}) \frac{w_{i,j}}{\phi_{i,j}} \left(y_{i,j} \frac{\mu_{i,j}^{1-p}}{1-p} - \frac{\mu_{i,j}^{2-p}}{2-p} \right).$$

Pour trouver p , en utilisant le principe du maximum de vraisemblance, un logiciel statistique tel que SAS peut optimiser tous les paramètres à la fois. On obtient alors les résultats du tableau 4.3. Le modèle utilisé ici suppose que les observations de la colonne 11 appartiennent à la colonne 10, et donc qu'ils ont le même facteur de surdispersion. Cette restriction est imposée dû au manque d'observations pour la dernière colonne. Par ailleurs, on suppose que l'exposition est le volume d'affaires tel que dans (Wüthrich, 2003). Curieusement, le facteur p optimal est assez différent selon la structure de la dispersion utilisée. D'une part, lorsque le paramètre de surdispersion est constant, on obtient $p = 1.1741$ alors qu'ici, on trouve $p = 1.8111$. On en déduit que p et $\phi_{i,j}$ sont très liés. Pour conclure l'analyse, les paramètres ainsi que leur écart-type sont disponibles dans le tableau 5.10 (p.137) en annexes. Les réserves et l'EQMP sont présentées dans la section 4.4, alors que le triangle de l'encouru cumulatif est illustré dans le tableau 5.9 (p.136).

4.2.2 Incorporation d'une tendance dans la variance du modèle

La structure du modèle nous permet d'introduire une tendance $f(j)$ dans le rapport des variances. Pour ce faire, on utilise une procédure analogue à celle de la courbe d'Hoerl (section 3.1.8). Par exemple, à titre démonstratif, supposons que nous avons une tendance linéaire dans le temps ($f(j) = j$) :

$$\log(\phi_{i,j}) = \log(\phi) + \gamma \log j ,$$

Période	1	2	3	4	5	6	7	8	9	10	11
ϕ_j	565	929	1 243	1 528	1 794	2 045	2 284	2 514	2 737	2 952	3 161

Tableau 4.4 ϕ_j selon la période, $p = 1.4157$, $\phi = 564.71$, $\gamma = 0.7184$

où γ est le paramètre qui représente la force de l'effet linéaire des périodes de développement et ϕ est le facteur de surdispersion de la cellule de référence. Si le paramètre γ n'est pas significativement différent de 0, alors il faut alors considérer une autre fonction de rapports de variance $f(j)$, ou bien peut être ajouter une tendance log-linéaire dans le temps (section 3.1.8). Ainsi, en optimisant en même temps les paramètres p , ϕ , γ et $\tilde{\mu}$, on obtient les résultats du tableau 4.4. Les paramètres $\tilde{\mu}$ sont très similaires à ceux du tableau 3.5 et ne sont pas présentés ici. De plus, puisque les résultats de l'encouru cumulatif sont très proches de ceux du modèle de la section précédente, il ne sont pas présentés.

On voit que la méthode par estimation directe ainsi que la méthode avec incorporation d'une tendance linéaire indiquent qu'il est préférable d'attribuer un facteur de surdispersion qui est plus grand à mesure que la période de développement augmente. L'ordre de grandeur des facteurs de surdispersion est différent d'un modèle à l'autre en raison de la valeur de leur paramètre p . Néanmoins, quoique instructifs sur la nature du risque, les méthodes basées sur le principe de vraisemblance maximale ont tendance à sous-estimer la valeur du paramètre de surdispersion, et donc il faut reconsidérer l'amplitude des paramètres de surdispersion lorsqu'on établit l'incertitude des réserves. La section 4.4 traite plus en profondeur ce concept.

4.3 Modèles linéaires généralisés doubles

Les modèles linéaires généralisés doubles² (*DGLM*) ont les mêmes objectifs que les modèles de dispersion. À l'origine, ces modèles sont apparus dans un contexte où, après avoir modélisé les moyennes selon différents facteurs explicatifs, on a observé que dans certains cas, la variabilité de certaines données autour de leur espérance est plus grande que leur variabilité théorique. De plus, le facteur de surdispersion était souvent différent selon les facteurs explicatifs. L'idée des *DGLM* est de modéliser la variance en même temps que la moyenne afin de trouver les facteurs de surdispersion qui reflètent le mieux la situation. Ceux-ci dépendent d'un nouveau vecteur de variables explicatives. Ainsi, chaque observation aura une moyenne et une variance qui dépendent de leur propre vecteur explicatif.

Ce concept a été introduit par (Smyth, 1989). Cette construction a été reprise par (Lee et Nelder, 2006) dans un contexte encore plus général. On appelle ces modèles *doubles* parce qu'ils utilisent deux modèles parallèles qui s'optimisent conjointement. L'un modélise les moyennes, l'autre la déviance, qui est une mesure de la variance. L'article de (Smyth et Jorgensen, 2002) présente un exemple concret d'un *DGLM* appliqué en assurance de dommages dans un contexte de tarification où on possède peu de paramètres et beaucoup de données. Le contexte est quelque peu différent de celui des réserves actuarielles où il y a peu d'observations et possiblement beaucoup de paramètres.

Cette méthode a été développée dans le but de rester dans un contexte de modèles linéaires généralisés. Ainsi, on peut retrouver les résultats asymptotiques classiques des estimateurs qui sont très utiles lors de l'inférence sur les paramètres et les données. Par exemple, nous pouvons alors utiliser une approximation des deux premiers moments

2. Double Generalized Linear Models (DGLM)

appelée aussi approximation du point de selle³, et tant que la moyenne et la dispersion sont bien spécifiées (Smyth et Jorgensen, 2002), les estimateurs convergent vers leur vraie valeur même si on s'est trompé de loi.

Enfin, les modèles linéaires généralisés doubles utilisent le principe de la déviance pour estimer tous les paramètres qui estiment le facteur de surdispersion. Ainsi, puisque les logiciels d'optimisation utilisent implicitement le principe de vraisemblance maximale, il faut trouver un algorithme différent pour incorporer le principe de la déviance.

Définition du modèle

On suit de près la méthodologie expliquée dans (Smyth et Jorgensen, 2002). Le modèle est construit en deux étapes. On modélise d'abord les pertes, et ensuite on observe l'écart des observations par rapport aux prédictions. Cet écart, appelé *résidu de la déviance* ($d_{i,j}$), est une mesure de la variabilité qui sert à identifier les facteurs qui contribuent le plus à la variance. La somme de ces résidus est appelée *déviance* (D) (section 3.1.9).

Il y a deux approximations lors de cette estimation. Premièrement, on suppose que $E[d_{i,j}] = \phi_{i,j}$. En d'autres mots, on suppose que l'espérance de la déviance est égale au facteur de surdispersion. En général, il n'y a qu'un biais léger, sauf dans les cas extrêmes, par exemple avec une loi de Poisson avec des moyennes petites (Nelder et al., 1998). Ces biais affectent l'analyse seulement au niveau des poids attribués aux données et ont un petit impact sur l'estimation des moyennes. La deuxième approximation est la supposition intrinsèque que la déviance suit une loi gamma, peu importe la loi choisie pour modéliser la moyenne. Cette procédure est utilisée car la déviance peut être approximée par une Chi-deux (loi gamma) même quand le modèle de la moyenne n'est pas une loi normale (Smyth, 1989).

3. Saddlepoint Approximation

La méthode présentée est appliquée à un modèle de Tweedie avec $p \in (1, 2)$. On suppose le coût total de chaque cellule dépend de l'année d'accident i et de la période de développement j comme dans les sections précédentes. En modélisant les coûts totaux par unité d'exposition $w_{i,j}$ et en reprenant l'équation (3.2), nous avons :

$$\mu_{i,j} = \exp\{X_{i,j}\tilde{\mu}\},$$

où $X_{i,j}$ est la ligne (i, j) de la matrice catégorielle X , qui indique quels paramètres du vecteur $\tilde{\mu}$ seront utilisés. Le vecteur $\tilde{\mu}$ contient tous les paramètres des lignes et des colonnes pour modéliser les moyennes des coûts totaux (section 3.1).

Ensuite, on observe les écarts du modèle avec les observations passées. On suppose que ces déviations peuvent être expliquées par l'année d'accident et la période de développement grâce à un modèle gamma utilisé parallèlement. La déviation prévue est notée $\phi_{i,j}$.

$$\phi_{i,j} = \exp\{V_{i,j}\tilde{\gamma}\},$$

où la matrice V est la matrice catégorielle des paramètres associés aux variances, et le vecteur $\tilde{\gamma}$ indique les paramètres de lignes et colonnes associés à la modélisation de la déviance (variance). La matrice V peut aussi être égale à X .

On trouve les vecteurs $\tilde{\mu}$ et $\tilde{\gamma}$ optimaux en alternance jusqu'à ce qu'il y ait convergence. À chaque itération, on suppose que le vecteur complémentaire est gardé fixe et que seule une partie du modèle est optimisée à la fois. Pour des fins d'identification, nous nommerons *modélisation des coûts* l'estimation des $\tilde{\mu}$ et *modélisation des variances* l'estimation des $\tilde{\gamma}$, même si en fait le tout fait partie d'un modèle double.

Puisque le principe de la déviance n'utilise pas le nombre de réclamations, on peut se contenter uniquement du triangle des coûts totaux. Cependant, la connaissance

du nombre de réclamations permet d'estimer p de la loi de Tweedie plus facilement et donne plus de précision au modèle en général.

On trouve d'abord les paramètres optimaux des moyennes selon une paramétrisation naturelle (section 3.1.7). En utilisant les mêmes hypothèses que dans la section 3.2.3, on modélise en tenant compte de l'exposition. Alors, on a :

$$Y_{i,j} \sim \text{Tweedie}(\mu_{i,j}, \frac{w_{i,j}}{\phi_{i,j}}, 1) \quad , \quad \mu_{i,j} = \exp\{X_{i,j}\tilde{\mu}\} \quad ,$$

où la moyenne de chaque observation dépend de la ligne i et de la colonne j selon un lien multiplicatif alors que l'exposition est égale à $w_{i,j}/\phi_{i,j}$ et le facteur de surdispersion est égal à 1. Aussi, $w_{i,j}$ représente l'exposition initiale, alors que $\phi_{i,j}$ sont les ajustements à l'exposition initiale qui, à la première itération, sont nuls ($\phi_{i,j} = 1$, $\forall i, \forall j$). Le calcul des ajustements subséquents est expliqué plus loin. Après avoir modélisé la surdispersion, les ajustements selon les nouveaux facteurs de surdispersion seront intégrés dans le modèle uniquement dans l'exposition. Donc, le facteur de surdispersion continue à être égal à 1 partout.

Ensuite, on trouve les résidus de la déviance de chaque cellule par rapport à la valeur observée. Les formules de ces résidus pour un modèle de Tweedie et pour un modèle gamma sont disponibles dans la figure 4.2. Comme présenté dans la section 3.1.9, avec un facteur de surdispersion égal à 1, on obtient :

$$d_{i,j} = 2w_{i,j} \left(y_{i,j} \frac{y_{i,j}^{1-p} - \mu_{i,j}^{1-p}}{1-p} - \frac{y_{i,j}^{2-p} - \mu_{i,j}^{2-p}}{2-p} \right) \quad .$$

Si les résidus de la déviance sont grands, alors le facteur de dispersion calculé subséquent sera plus élevé. Ensuite, on applique le deuxième sous-modèle :

$$d_{i,j} \sim \Gamma(\phi_{i,j}, 2) \quad , \quad \phi_{i,j} = \exp\{V_{i,j}\tilde{\gamma}\} \quad ,$$

où V est la matrice catégorielle des facteurs de surdispersion et où la loi gamma est paramétrisée de sorte à ce que :

$$E[d_{i,j}] = \phi_{i,j} \quad , \quad \text{Var}[d_{i,j}] = 2\phi_{i,j}^2 \quad .$$

Composante	Modélisation des coûts	Modélisation de la déviance
Variable à modéliser	Y^*	d
Moyenne	λ	ϕ
Variance	$\phi V(\lambda)$	$2\phi^2$
Lien	$\eta = \eta(\lambda)$	$\eta_d = g_d(\phi)$
Prédicteur linéaire	$\eta = \mathbf{X}\beta$	$\eta_d = \mathbf{V}\gamma$
Déviance	$2 \int_{\lambda}^{\infty} \frac{\eta - \lambda}{V(\lambda)} d\lambda$	$2(-\log(d/\phi)) + (d - \phi)/\phi$
Poids	w/ϕ	$\frac{1}{\phi}$

Figure 4.2 Aperçu des interrelations entre les deux sous-modèles

Le nombre de paramètres associés aux variances peut être différent de celui pour les moyennes. Par exemple, on peut supposer que les facteurs affectant la variance sont seulement les périodes de développement, auquel cas, le vecteur $\tilde{\gamma}$ possédera moins de paramètres que le vecteur $\tilde{\mu}$. Le deuxième sous-modèle utilise une loi gamma avec moyenne ϕ qui dépend de la matrice catégorielle V selon une fonction de lien g_d (ici c'est un lien multiplicatif aussi), avec une exposition uniforme (1 partout) et un facteur de dispersion de 2. La figure 4.2 présente les relations entre les deux sous-modèles⁴.

Ensuite, après avoir estimé le vecteur $\tilde{\gamma}$, nous avons des paramètres de surdispersion prévus $\tilde{\phi} = \{\phi_{1,1}, \phi_{2,1}, \dots, \phi_{1,J}\}$ qui sont possiblement différents d'une observation à l'autre. Rappelons qu'un facteur de dispersion élevé implique qu'on a moins confiance envers la donnée à laquelle il est associé car la variabilité du modèle associée à cette observations est plus grande. L'idée est de donner moins d'importance aux données avec une grande variabilité lors de la prochaine estimation des moyennes. La façon la plus simple de mettre sur pied cette relation est de diviser l'exposition initiale par le facteur de dispersion pour obtenir la nouvelle exposition, notée w' , qui sera utilisée dans la modélisation de la moyenne lors de l'itération suivante. Cette approche se justifie intuitivement par le fait que l'estimation des moyennes dépend de la relative importance entre leur exposition. En d'autres mots, savoir qu'une observation a une exposition deux fois plus grande qu'une autre aura un impact sur les moyennes estimées. Par contre,

4. L'image est inspirée de (Nelder et al., 1998)

multiplier l'exposition par deux pour toutes les observations fera changer la valeur de la fonction de vraisemblance, mais n'aura pas d'impact sur l'estimation des moyennes.

Après avoir mis à jour les nouveaux poids, on recommence la boucle en réestimant les pertes moyennes. Ensuite, on continue les mêmes étapes jusqu'à ce qu'il y ait convergence. Généralement, la convergence s'effectue rapidement, car les deux sous-modèles sont indépendants en ce qui a trait à l'optimisation par les dérivées de la log-vraisemblance (Smyth, 1989).

Pour résoudre parallèlement ces deux sous-modèles, on utilise généralement une adaptation de la méthode de scores de Fisher. D'abord, on choisit des valeurs raisonnables de départ pour les vecteurs $\tilde{\mu}$ et $\tilde{\gamma}$. Ensuite, à chaque itération, on met à jour d'abord le vecteur $\tilde{\mu}$ et ensuite le vecteur $\tilde{\gamma}$ en utilisant la méthode des moindres carrés pondérés⁵.

Pour trouver la $(k + 1)^e$ itération de $\tilde{\mu}$, on utilise la formule suivante :

$$\begin{aligned} \tilde{\mu} & \text{ calculé à partir d'une loi Tweedie,} \\ \tilde{\mu}^{k+1} &= (X^T W X)^{-1} X^T W z, \end{aligned}$$

où $\tilde{\mu}^{k+1}$ est fonction de l'itération précédente $\tilde{\mu}^k$ et du vecteur associé aux facteurs de surdispersion $\tilde{\gamma}$. La loi de Tweedie a un paramètre p et une exposition initiale w , l'information est représentée en vecteurs. Aussi, W est une matrice diagonale avec les éléments suivants :

$$W = \text{diag} \left(\left(\frac{\partial g(\mu)}{\partial \mu} \right)^{-2} \frac{w}{\phi V(\mu)} \right), \quad (4.2)$$

où la fonction de variance est $V(\mu) = \mu^p$, et z est égal à :

$$z = \frac{\partial g(\mu)}{\partial \mu} (y - \mu) + g(\mu),$$

5. Iterative Weighted Least Squares (IWLS)

et finalement X est la matrice catégorielle pour les coûts, alors que μ est le vecteur des moyennes des coûts prévus calculé comme suit :

$$\mu = \exp\{X\tilde{\mu}^{(k)}\} ,$$

et w est le vecteur des expositions qui est associé aux observations et $g(\cdot)$ est la fonction de lien. Pour un lien multiplicatif, $g(\mu) = \log(\mu)$.

Pour $\tilde{\gamma}$, on a :

$$\tilde{\gamma} \quad \text{calculé à partir d'une loi gamma ,} \quad (4.3)$$

$$\tilde{\gamma}^{k+1} = (V^T W_d V)^{-1} V^T W_d z_d , \quad (4.4)$$

où $\tilde{\gamma}^{k+1}$ est fonction de l'itération précédente $\tilde{\gamma}^k$, et W_d est une matrice diagonale avec les éléments :

$$W_d = \text{diag} \left(\left(\frac{\partial g_d(\phi)}{\partial \phi} \right)^{-2} \frac{1}{2V(\phi)} \right) , \quad (4.5)$$

où la fonction de variance est $V(\phi) = \phi^2$, et z_d est égal à :

$$z_d = \frac{\partial g_d(\phi)}{\partial \phi} (d - \phi) + g_d(\phi) , \quad (4.6)$$

et V est la matrice catégorielle pour les variances. Le vecteur ϕ est calculé à partir de $\tilde{\gamma}^{(k)}$:

$$\phi = \exp\{V\tilde{\gamma}^{(k)}\} .$$

La procédure des moindres carrés pondérés est une généralisation des moindres carrés. Rappelons que ceux-ci sont utilisés dans un modèle linéaire Normal classique (section 1.7.5) et ont pour solution :

$$\tilde{\mu} = (X^T X)^{-1} X^T Y ,$$

où X est la matrice catégorielle et Y est le vecteur des observations. Les modèles de la famille exponentielle linéaire traités dans cet ouvrage peuvent s'exprimer à l'aide d'une

fonction de variance sans avoir besoin de définir la fonction de densité (section 3.2.3). Ainsi, puisque seulement les deux premiers moments doivent être connus pour connaître ce type de distribution, on peut exprimer les données par une approximation linéaire et quadratique et obtenir la méthode des moindres carrés pondérés.

Approximation avec déviance restreinte (REML)

La littérature a souvent indiqué que les estimateurs issus du principe de vraisemblance maximale sont biaisés vers le bas lorsque le nombre de paramètres est grand comparativement au nombre d'observations. La même remarque s'applique pour la méthode des moindres carrés pondérés. On utilise alors la correction basée sur la déviance restreinte⁶. Soit h les éléments de la diagonale de la matrice :

$$h = W^{1/2}X(X^TWX)^{-1}X^TW^{1/2}.$$

(Smyth et Jorgensen, 2002) propose d'ajuster les équations (4.4), (4.5) et (4.6) de la façon suivante :

$$\tilde{\gamma}^{k+1} = (V^TW_d^*V)^{-1}V^TW_d^*z_d^*,$$

où W_d^* est la matrice diagonale :

$$\begin{aligned} W_d^* &= \text{diag} \left(\left[\frac{\partial g_d(\phi)}{\partial \phi} \right]^{-2} \frac{1-h}{2V(\phi)} \right), \\ z_d^* &= \frac{\partial g_d(\phi)}{\partial \phi} \left(\frac{d}{1-h} - \phi \right) + g_d(\phi). \end{aligned}$$

La procédure des moindres carrés pondérés ajustés décrite ci-dessus optimise approximativement la fonction de vraisemblance restreinte suivante :

$$l^*(y, \tilde{\mu}, \tilde{\gamma}, p) = l(y, \tilde{\mu}, \tilde{\gamma}, p) + \frac{1}{2} \log |X^TWX|, \quad (4.7)$$

6. Restricted or Residual Maximum Likelihood (REML)

où $l(y, \tilde{\mu}, \tilde{\gamma}, p)$ est la fonction de vraisemblance de l'équation (3.17) (p.77). W est la matrice telle que définie dans l'équation (4.2) et X est la matrice catégorielle associée aux observations y . En d'autres mots, on peut soit utiliser la procédure des moindres carrés pondérés ajustés, soit directement chercher à optimiser l'équation (4.7) via le principe de vraisemblance maximale. On peut consulter (Smyth et Jorgensen, 2002) ainsi que (Smyth, Huele et Verbyla, 2001) pour une discussion approfondie de ces concepts.

Trace de l'algorithme avec p fixé

Pour bien représenter le contexte des triangles de développement, on utilise les déviations restreintes (REML), et les paramètres des facteurs de surdispersion sont associés aux colonnes seulement. L'algorithme DGLM avec approximation avec la déviance restreinte a été appliqué aux données de la *Swiss Motor Insurance* (p. 17).

Le manque d'observations occasionne quelques problèmes pratiques qui méritent notre attention. Tout d'abord, il faut se rappeler que la dernière colonne ne possède qu'une observation et donc, puisque le modèle possède un paramètre associé à cette colonne exclusivement, la valeur prévue est égale à la valeur observée. Dans ces conditions, il est difficile d'associer une notion de modélisation de la variance car il manque des degrés de liberté. Pour faire face à ce problème, deux ajustements sont faits. D'abord, lors de la paramétrisation de V , le même paramètre représente les quatre dernières colonnes, comme illustré dans le tableau 4.5. Ensuite, l'observation de la dernière colonne est ignorée lors des calculs de la déviance. Pour ce faire, pour cette observation en particulier, on fixe la déviance à 1, et les expositions ajustées à chaque itération restent inchangées. En général, l'algorithme diverge lorsqu'on ne fait pas le dernier ajustement.

Colonne	1	2	3	4	5	6	7	8	9	10	11
Paramètre	γ_0	$\gamma_0 + \gamma_1$	$\gamma_0 + \gamma_2$	$\gamma_0 + \gamma_3$	$\gamma_0 + \gamma_4$	$\gamma_0 + \gamma_5$	$\gamma_0 + \gamma_6$	$\gamma_0 + \gamma_7$	$\gamma_0 + \gamma_7$	$\gamma_0 + \gamma_7$	$\gamma_0 + \gamma_7$

Tableau 4.5 Paramétrisation de V , la matrice catégorielle associée aux facteurs de surdispersion

Les figures 4.3, 4.4 et 4.5 (p. 117) donnent un aperçu des résultats intermédiaires lorsque p est connu et fixé à 1.5. Dès la 2^e itération, on obtient des valeurs pour l'exposition et la surdispersion très proches des résultats finaux. Cependant, après la première itération très fructueuse, les ajustement subséquents sont très modérés. Dépendamment du degré de précision souhaité, on peut nécessiter au-delà de 100 itérations dans certains cas.

Par ailleurs, on peut s'intéresser au nombre de colonnes explicatives pour modéliser la variance. Tout d'abord, il faut d'emblée enlever le paramètre associé à la colonne ultime car le modèle prévoit exactement la valeur observée. Ensuite, l'algorithme a été testé avec 10 paramètres pour la surdispersion. En fin de compte, les résultats n'ont pas été concluants ; l'algorithme diverge. Il a fallu réduire le nombre de dimensions à 7 pour obtenir une convergence. On en conclut que les DGLMs nécessitent un minimum de données pour pouvoir opérer.

Similairement, on a essayé la procédure inverse et d'ajouter les paramètres un à un. En commençant avec un seul paramètre, l'algorithme converge pour chaque p , mais diverge lorsqu'on essaie d'optimiser p (section 3.2.3). En essayant avec deux paramètres, il diverge aussi pour certaines valeurs de p . Pour l'instant, il est difficile d'isoler les conditions qui font en sorte que la convergence s'effectue ou non. L'étude de ce phénomène peut être l'objet de recherches futures.

Optimisation de p

L'optimisation du paramètre p dans un contexte de DGLMs est un sujet délicat. (Smyth et Jorgensen, 2002) optimise p en utilisant le principe de la déviance comme dans la section 3.2.3.

Puisque la méthode des moindres carrés pondérés ne possède pas une étape particulière pour optimiser p , on utilise le principe de vraisemblance maximale conjointement au principe de la déviance basé sur la méthode des moindres carrés pondérés. Ainsi, lorsqu'on optimise les paramètres $\tilde{\mu}$ et $\tilde{\gamma}$, on utilise les moindres carrés pondérés. Lorsqu'on optimise le paramètre p , on utilise le principe de vraisemblance en utilisant la fonction 4.7.

L'approche consiste à optimiser p en même temps que tous les autres paramètres. Pour ce faire, on procède de la façon suivante :

1. On établit une étendue de valeurs p à vérifier. Commencer par $p \in \{1.1, 1.2, \dots, 1.9\}$
2. Dans chaque cas, on trouve $\tilde{\mu}$ et $\tilde{\gamma}$ optimaux en gardant p fixe
3. On compare les fonctions de vraisemblance, et on recommence l'étape 1 en utilisant une étendue plus petite qui est autour du point avec la meilleure vraisemblance

Ici, on suppose que les fonctions de vraisemblance sont comparables malgré des paramètres associés aux variances p et $\tilde{\gamma}$ différents. En ce qui concerne p , on peut se convaincre que la comparaison est justifiable en utilisant les mêmes arguments que la section 3.2.3. Néanmoins, en appliquant tout de même cet algorithme, on obtient la figure 4.6 (p. 120).

On obtient une valeur optimale $p = 1.6625$ ainsi que les valeurs du tableau 4.6. On constate que les facteurs de surdispersion ont un ordre de grandeur semblable au modèle avec le principe de déviance à un seul paramètre.

Période	1	2	3	4	5	6	7	8	9	10	11
ϕ											
Deviance 1 paramètre	25 169	25 169	25 169	25 169	25 169	25 169	25 169	25 169	25 169	25 169	25 169
DGLM 8 paramètres	5 834	490	8 757	9 413	45 502	18 382	50 551	78 261	78 261	78 261	78 261

Tableau 4.6 Facteurs de surdispersion selon un modèle de Tweedie, $p = 1.6625$

Pour trouver le facteur de surdispersion correspondant à chaque colonne, on fait le calcul suivant :

$$\phi_{i,j} = \frac{w_{i,j}}{w'_{i,j}}.$$

Donc, le facteur de surdispersion est ce qui transforme l'exposition initiale en exposition finale lorsque le modèle a convergé.

Par ailleurs, les pertes espérées sont égales aux pertes prévues pour l'observation de la colonne 11. Donc, la déviance sera automatiquement nulle. Alors, pour pallier à ce problème, on modélise l'ensemble des observations restantes uniquement, et à la fin, on suppose artificiellement que le facteur de surdispersion est le même que celui associé à la colonne précédente.

Le triangle de l'encours cumulatif ainsi que les paramètres du DGLM avec $p = 1.6625$ sont disponibles dans les tableaux 5.11 (p. 138) et 5.12 (p. 139) respectivement.

4.4 Comparaisons entre les modèles de dispersion et les doubles GLMs

4.4.1 Intégration des facteurs de surdispersion dans le calcul de l'EQMP

Si on utilise le principe de vraisemblance maximale tel que dans les modèles de dispersion, alors en reprenant l'équation (3.23), on peut montrer que l'EQMP se calcule

de la façon suivante :

$$\begin{aligned} \text{EQMP}[\hat{R}_i] &= \sum_{(i,j) \in L} \phi_{i,j} w_{i,j} \mu_{i,j}^p + \sum_{(i,j) \in L} (w_{i,j} \mu_{i,j})^2 \text{Var}[\eta_{i,j}] \\ &+ \sum_{(i,j) \in L, (i_1,j_1) \neq (i_2,j_2)} w_{i,j_2} \mu_{i,j_2} \text{Cov}(\eta_{i,j_1}, \eta_{i,j_2}) \end{aligned} \quad (4.8)$$

où L représente les observations futures de l'année d'accident i . Par ailleurs, on utilise la matrice de variance-covariance tronquée, c'est-à-dire on ne garde que les variances et covariances des paramètres relatifs aux moyennes. On ne multiplie pas cette matrice par un facteur de surdispersion car elle est naturellement ajustée lors de l'optimisation. En d'autres mots, lorsque les paramètres de surdispersion sont estimés, alors les valeurs estimées de la matrice de variance-covariance sont plus grandes que dans pour un modèle avec ϕ fixé.

Pour les DGLM, en partant encore une fois de l'équation (3.23), on en déduit la formule suivante :

$$\begin{aligned} \text{EQMP}[\hat{R}_i] &= \sum_{(i,j) \in L} \phi_{i,j} w_{i,j} \mu_{i,j}^p + \sum_{(i,j) \in L} \phi_{i,j} (w_{i,j} \mu_{i,j})^2 \text{Var}[\eta_{i,j}] \\ &+ \sum_{(i,j) \in L, (i_1,j_1) \neq (i_2,j_2)} \sqrt{\phi_{i,j_1} \phi_{i,j_2}} w_{i,j_1} \mu_{i,j_1} w_{i,j_2} \mu_{i,j_2} \text{Cov}(\eta_{i,j_1}, \eta_{i,j_2}) \end{aligned} \quad (4.9)$$

où $\phi_{i,j}$ est le facteur de surdispersion qui dépend des lignes et colonnes. La matrice de variance-covariance est celle issue de la dernière itération des paramètres relatifs aux moyennes. Puisqu'ici, les paramètres de surdispersion sont fixés, il faut multiplier la matrice de variance-covariance par les facteurs de surdispersion $\phi_{i,j}$ qui lui correspondent. Dans tous les cas, on utilise les expositions initiales $w_{i,j}$ et non l'exposition issue du modèle qui a convergé.

Par ailleurs, il s'avère que la variabilité supplémentaire suite à l'estimation de p n'affecte pas la variabilité des paramètres des moyennes car ils sont orthogonaux en ce

Obs.	Principe	1	2	3	4	5	6	7	8	9	10	11
0	Déviante	3 921	3 921	3 921	3 921	3 921	3 921	3 921	3 921	3 921	3 921	3 921
0	V. 1 paramètre	199	199	199	199	199	199	199	199	199	199	199
0	V. 9 paramètres	202	169	259	384	716	648	1 003	1 269	901	904	904
7	Déviante	1 857	1 857	1 857	1 857	1 857	1 857	1 857	1 857	1 857	1 857	1 857
7	V. 1 paramètre	198	198	198	198	198	198	198	198	198	198	198
7	V. 9 paramètres	202	166	250	384	637	648	800	964	898	903	903
14	Déviante	1 117	1 117	1 117	1 117	1 117	1 117	1 117	1 117	1 117	1 117	1 117
14	V. 1 paramètre	198	198	198	198	198	198	198	198	198	198	198
14	V. 9 paramètres	204	166	250	385	637	648	922	969	880	904	904

Tableau 4.7 Facteur de surdispersion en fonction du principe, de la colonne et du nombre d'observations ajustées, pour un modèle où $p = 1.1741$ et $w = 1$

qui a trait à l'optimisation de la fonction vraisemblance (Smyth et Jorgensen, 2002), et donc ne sont pas tenus en compte dans les équations (4.8) et (4.9).

4.4.2 Analyse de sensibilité du paramètre de surdispersion issu du principe de vraisemblance maximale

La valeur du paramètre de surdispersion moyenne peut différer significativement selon le principe utilisé lorsque le nombre de données est petit par rapport aux nombre de paramètres estimés. En effet, le tableau 4.7 indique que le paramètre de surdispersion calculé selon le principe de la déviance est largement supérieur à celui selon le principe de vraisemblance maximale. Pour les trois modèles du tableau 4.7, on suppose que $p = 1.1741$ afin de mieux comparer les facteurs de surdispersion.

La vraisemblance à un paramètre est associée à la situation où il n'y a qu'un facteur de surdispersion pour toutes les observations. Cette différence de grandeur a un impact très significatif sur l'EQMP. Il s'avère qu'en général, le calcul selon le principe de la déviance ou de Pearson retourne des valeurs plus grandes que le principe de vraisemblance maximale. Ceci est dû à leur grande sensibilité aux déviations extrêmes lorsque la base de données est petite (Ruoyan, 2004). La colonne *Obs.* indique le nombre d'observations modifiées pour qu'elles coïncident avec leur moyenne théorique. Ainsi, lorsqu'on

Cellule		Principe	
<i>i</i>	<i>j</i>	Maximum de vraisemblance	Déviance
1	1	14,19	149,56
2	1	15,28	0,77
3	1	15,58	0,02
4	1	15,14	8,32
5	1	14,49	3,31
6	1	14,84	30,41
7	1	13,73	4,93
8	1	13,38	11,30
9	1	13,52	31,57
1	2	6,90	267,52
2	2	6,28	11,70
3	2	6,03	10,56
2	10	0,06	4,34
1	11	0,00	0,00
Total		199,28	3 921,00

Tableau 4.8 Contribution au facteur de surdispersion de chaque observation

change les 14 observations les plus extrêmes, le facteur de surdispersion est relativement pareil pour les deux principes, alors que le facteur de surdispersion issu du principe de la déviance chute notablement.

Le tableau 4.8 illustre la contribution de chaque observation au facteur de surdispersion selon le principe de la déviance ou du maximum de vraisemblance. On découvre que la contribution provient essentiellement des données avec un nombre de réclamations élevé pour la vraisemblance. À l'opposé, la contribution à la déviance provient en grande partie des observations qui s'éloignent de leur prédiction. Il s'avère que malgré cette différence marquée, lorsque l'ensemble de données est assez grand, la contribution totale devient semblable pour les deux principes (Ruoyan, 2004).

Ainsi, le choix du principe a un impact important sur le calcul de l'EQMP. Puisque les triangles de développement possèdent relativement peu d'observations et beaucoup de paramètres, la situation exige qu'on fasse attention dans le choix du principe et à l'interprétation des résultats.

Année	$w_{i,j}$	Déviance		Dispersion 9 paramètres		DGLM 9 paramètres	
		R_i	EQMP ^{1/2}	R_i	EQMP ^{1/2}	R_i	EQMP ^{1/2}
1	112 953	-	-	-	-	-	-
2	110 364	326	2 638	324	736	291	2 809
3	105 400	21 565	26 804	21 352	27 769	19 586	32 040
4	102 067	40 716	35 556	40 185	34 658	37 191	41 717
5	99 124	89 298	53 297	87 224	54 103	83 592	68 788
6	101 460	138 335	66 052	138 203	62 649	131 053	80 811
7	94 753	204 262	80 906	202 469	69 443	206 128	91 283
8	92 326	360 484	111 999	359 148	82 510	343 778	137 516
9	89 545	597 056	150 003	596 119	93 407	560 207	151 951
Total	-	1 452 042	271 886	1 445 024	234 818	1 381 827	334 269

Tableau 4.9 Réserve et EQMP du modèle de Tweedie selon différentes approches pour la surdispersion

En pratique, on applique le principe de la déviance pour calculer l'EQMP comme par exemple dans (Wüthrich, 2003) et (Wüthrich et Merz, 2008). De plus, puisque le facteur de surdispersion est généralement plus élevé avec le principe de la déviance, les réserves obtenues sont plus conservatrices.

4.4.3 Analyse des résultats

Les tableaux 4.9 et 4.10 montrent les réserves espérées et l'EQMP selon les différents modèles. On constate que puisque le principe de vraisemblance maximale propose un facteur de surdispersion plus petit, l'EQMP sera aussi plus petite. Néanmoins, les trois approches proposent des réserves espérées très semblables.

4.5 Discussion et recherche future

Tout d'abord, quelques commentaires concernant les modèles linéaires généralisés doubles. Puisque les estimateurs du maximum de vraisemblance ont tendance à sous-estimer la dispersion lorsque le nombre de paramètres est grand par rapport au nombre d'observations, on peut penser à utiliser une fonction de vraisemblance qui tient compte de cette imprécision. Comme pour les DGLM, on peut penser à utiliser la fonction de

Année	Déviance $p = 1.1741$			Dispersion 9 paramètres $p = 1.8111$			DGLM 9 paramètres $p = 1.6625$		
	Estimation	Statistique	EQMP ^{1/2}	Estimation	Statistique	EQMP ^{1/2}	Estimation	Statistique	EQMP ^{1/2}
1	-	-	-	-	-	-	-	-	-
2	1 869	1 861	2 638	546	493	736	2 728	668	2 809
3	15 601	21 795	26 804	16 984	21 970	27 769	23 141	22 161	32 040
4	19 144	29 962	35 556	20 001	28 304	34 658	28 436	30 524	41 717
5	25 976	46 538	53 297	28 127	46 217	54 103	39 518	56 304	68 788
6	30 564	58 556	66 052	32 881	53 327	62 649	44 424	67 505	80 811
7	35 230	72 833	80 906	34 783	60 103	69 443	48 145	77 554	91 283
8	45 664	102 268	111 999	40 846	71 690	82 510	56 744	125 263	137 516
9	61 307	136 903	150 003	47 079	80 675	93 407	61 431	138 980	151 951
Total	180 126	203 658	271 886	183 345	146 711	234 818	248 122	223 989	334 269

Tableau 4.10 Décomposition de l'EQMP selon divers modèles

vraisemblance restreinte telle que définie dans la section 4.3. Par ailleurs, les modèles linéaires généralisés doubles peuvent aussi intégrer une tendance tant au niveau des paramètres correspondant aux moyennes que les paramètres correspondant à la surdispersion. La construction d'un tel modèle mérite d'être explorée.

D'autre part, il est intéressant de mentionner que Nelder et Lee, deux pionniers de l'élaboration de modèles statistiques, ont récemment établi les fondations d'un modèle encore plus global que les DGLMs. En effet, leur article (Lee et Nelder, 2006) introduit les *modèles hiérarchiques linéaires généralisés doubles*⁷ qui combine les modèles hiérarchiques⁸, avec les DGLM. L'exploration de cette nouvelle construction permet une flexibilité du modèle jusqu'à ce jour inégalée. Il est clair que ces outils statistiques possèdent un potentiel qui ne demande qu'à être exploité.

La modélisation à l'aide du principe de vraisemblance permet de construire des modèles très flexibles. Dans ce mémoire, il a été démontré comment intégrer l'exposition dans le modèle, mais on a omis de présenter un autre outil qui a un effet similaire : les

7. Double Hierarchical Generalized Lineral Model (DHGLM)

8. Hierarchical Models

*poids*⁹. Cette unité peut être associée à chaque observation individuellement et peut être interprétée comme le nombre d'observations indépendantes qui ont exactement la valeur observée pour la cellule donnée. Cette approche nous permet de jouer avec l'importance des données avant même que le modèle soit optimisé. De plus, ce concept peut être utilisé dans le cadre des DGLMs à la place de l'exposition pour avoir des variables qui sont plus facilement interprétables.

Dans un autre ordre d'idées, concernant le domaine des réserves stochastiques en général, un éventail de difficultés attendent d'être résolues. D'abord, il est difficile de quantifier l'erreur de modélisation. En général, on compare les résultats de plusieurs modèles et on sélectionne le meilleur modèle à partir du jugement. Néanmoins, on pourrait penser à développer des statistiques de vraisemblance entre les modèles, qui pourraient être basées par exemple sur les résidus. En ce qui concerne le modèle de Tweedie, on peut évaluer la variance du paramètre p grâce à la borne de Cramér-Rao, et ainsi possiblement construire un test qui permettrait de détecter si la différence entre le modèle de Tweedie et une de ses deux bornes, à savoir Poisson ou gamma, est significative. Finalement, il serait pertinent de développer des tests pour valider les hypothèses sous-jacentes à chaque modèle. L'article de (Peters, Shevchenko et Wüthrich, 2009) s'est penché sur la question et propose quelques approches qui tiennent compte de l'erreur de modélisation.

Ensuite, on peut penser à réduire le nombre de paramètres de chaque ligne et colonne par exemple en utilisant le test du Chi-deux. (Gluck et Venter, 2009) propose ainsi d'utiliser ces nouveaux degrés de liberté pour incorporer une tendance par année de calendrier (diagonales) afin de mieux capter l'effet de l'inflation. Une autre classe de modèles importante est celle basée sur l'analyse temporelle des données. Le filtre de Kalman en est un bon exemple. L'utilisation de séries chronologiques est aussi une

9. Weights

avenue qui est présentement étudiée dans (Gluck et Venter, 2009).

Par ailleurs, les modèles présentés ici ont pour la plupart que des domaines positifs. Hélas, il arrive qu'en pratique, la compagnie se fasse rembourser des paiements lors des périodes mures, enregistrant ainsi des paiements négatifs. Le modèle de Poisson surdispersé ne peut être appliqué que lorsque la somme des pertes incrémentales de chaque ligne et colonne est supérieure à 0 (Renshaw et Verrall, 1998). Il arrive que cette condition ne soit pas satisfaite, particulièrement dans les lignes d'affaires à cadence lente.

Pour faire face à ce problème, on peut penser à fixer les données négatives à 0. Alors, il faudra se rappeler que le modèle est biaisé et surestime les réserves. On peut aussi penser à utiliser des lois décalées. On translate toutes les données afin qu'elles soient positives, et on modélise normalement à partir de là. Finalement, on reconvertit les données à la fin. Par ailleurs, une autre idée serait de modéliser les valeurs cumulatives au lieu des valeurs incrémentales. Les valeurs cumulatives n'ont presque aucune chance d'être négatives. Cependant, le modèle est complètement différent, et l'interprétation des paramètres est sensiblement différente, comme le mentionne (Verrall, 2000).

(Kunkler, 1998) a effectué un survol complet des techniques pour traiter les données négatives. Il propose en plus de modéliser les données négatives parallèlement aux données positives, et de combiner les deux modèles ensemble à la fin. Cette approche est certainement très prometteuse, mais la validité statistique de cette approche soulève des questions.

Par ailleurs, (Schiegl, 2009) propose de modéliser la période entre la déclaration du sinistre et les paiements, parallèlement à la période entre la survenance du sinistre et la déclaration de celui-ci. Ces analyses s'effectuent alors dans un triangle en trois dimensions (pyramide).

Dans un autre ordre d'idées, la venue de la théorie statistique permet de modéliser la dépendance entre plusieurs lignes d'affaires. (Kirschner, Kerley et Isaacs, 2002) propose une généralisation de la technique de bootstrap pour modéliser implicitement cette dépendance. Hélas, la technique proposée ne permet pas de quantifier exactement l'amplitude de la dépendance, et quelles sont les données qui y contribuent. Aussi, (Brehm, 2002) propose une généralisation du modèle de Mack pour intégrer des coefficients de corrélation entre les lignes d'affaires. Cependant, peu de gens se sont penchés sur ce modèle nouveau et donc on ne connaît pas encore l'amplitude des bienfaits et des inconvénients de celui-ci. Par ailleurs, une avenue très prometteuse consiste à utiliser des copules pour modéliser cette dépendance.

À l'avenir, on peut penser à développer plus en profondeur ces nouveaux modèles, ou bien penser à utiliser des alternatives. Une idée très prometteuse est d'effectuer une analyse séparée de la fréquence et de la sévérité. Ensuite, on combine les résultats des deux analyses pour avoir les statistiques des coûts totaux. Cette approche permet d'intégrer le modèle de notre choix pour la fréquence et la sévérité, et donc nous ne sommes pas limités à un modèle somme composée Poisson-gamma tel que pour la loi de Tweedie.

Une autre approche est de structurer les pertes dans un fichier sinistres et de modéliser la perte ainsi que la période de développement, à partir de facteurs explicatifs propres à chaque assuré, comme proposé dans (Guszcza et Lommele, 2006). Cette méthode utilise davantage d'information, et permet de modéliser un portefeuille hétérogène. (Valdez, Antonio et Frees, 2009) est présentement en train de développer un modèle par fichier sinistre qui tient compte également du type de sinistre et de la dépendance entre ceux-ci.

Y	w	$\tilde{\mu}$	\tilde{d}	$\tilde{\gamma}$	$\tilde{\phi}$	w'
157.9516	112953	8.1402	8023.14	8.9209	7486.64	15.0573
176.8613	110364	0.0389	20.90	-0.1180	7486.64	14.7415
189.6096	105400	0.0941	174.40	0.6090	7486.64	14.0784
189.1466	102067	0.0786	788.06	0.5351	7486.64	13.6332
184.5313	99124	0.1513	7317.54	1.8658	7486.64	13.2401
185.6152	101460	0.0186	5812.75	0.6102	7486.64	13.5521
181.0255	94753	0.0924	1472.64	1.6176	7486.64	12.0503
179.9008	92326	0.0182	1461.54	2.0454	7486.64	12.3321
188.0111	89545	0.0332	4883.20		7486.64	11.9406
65.8896	112953	-1.1148	24264.63		6649.47	16.9568
60.3142	110364	-3.2262	1006.36		6649.47	16.6974
60.0330	105400	-4.2244	493.56		6649.47	15.8609
57.7051	102067	-4.5096	1834.47		6649.47	15.3497
58.4448	99124	-5.4992	8940.68		6649.47	14.9071
56.5860	101460	-5.8997	61.66		6649.47	15.2654
62.3546	94753	-5.9180	169.59		6649.47	14.2497
55.3601	92326	-6.8392	605.37		6649.47	13.8847
55.6590	89545	-6.8454	866.06		6649.47	13.4665
7.9273	112953	-11.0035	7155.15		15060.68	7.4999
8.8305	110364		11009.03		15060.68	7.3260
10.4361	105400		37467.41		15060.68	6.9984
7.7696	102067		756.76		15060.68	6.7771
6.9554	99124		4545.28		15060.68	6.5616
5.7343	101460		8981.01		15060.68	6.7367
5.5387	94753		20199.47		15060.68	6.2914
5.9926	92326		4930.63		15060.68	6.1303
5.4649	89545		13620.15		15060.68	5.9456
3.6099	112953		29124.70		12764.02	8.8355
1.4080	110364		49519.06		12764.02	8.6330
2.6532	105400		199.90		12764.02	6.2447
3.0276	102067		2296.35		12764.02	7.9640
2.9118	99124		0.46		12764.02	7.7537
2.4499	101460		233.84		12764.02	7.9365
0.2341	110364		3108.56		56060.01	1.9455
0.0028	112953		2451.47		56060.01	1.9009

Figure 4.3 Première itération du DGLM

Y	w	$\tilde{\mu}$	d	$\tilde{\gamma}$	$\tilde{\sigma}$	w'
157.9516	15.0873	0.1734	2.3434	-0.0255	0.91	16.5986
176.8613	14.7415	0.0311	0.0001	-0.2010	0.91	16.2181
189.6696	14.0784	0.0709	0.0002	-0.0193	0.91	15.4887
189.1468	13.8332	0.0502	0.0671	0.0057	0.91	14.9989
184.5313	13.2401	0.0554	0.0220	0.1296	0.91	14.5664
185.6152	13.5521	0.0097	0.3039	0.1481	0.91	14.0097
181.0255	12.8563	0.0352	0.0171	0.1724	0.91	13.9241
179.9608	12.3321	-0.0100	0.1479	0.1050	0.91	13.5674
168.0111	11.9606	0.0116	0.4344		0.91	13.1567
65.8896	16.9868	-1.1152	2.3208		0.74	22.8479
60.3142	16.5974	-3.2208	0.2684		0.74	22.3242
60.0330	16.8609	-4.7241	0.1438		0.74	21.3201
57.7061	15.3497	-4.5621	0.3285		0.74	20.6459
68.4448	14.9071	-5.4984	0.2371		0.74	20.0506
56.5860	15.2564	-5.8867	0.1188		0.74	20.5231
62.3546	14.2497	-5.9252	0.1743		0.74	19.1664
65.9601	13.8847	-6.8460	0.1181		0.74	18.6755
65.8890	13.4685	-6.8575	0.2226		0.74	18.1130
7.9273	7.4999	-11.0307	0.2024		0.69	6.4115
8.5305	7.3280		0.7724		0.69	8.2187
10.4381	6.9994		2.3468		0.69	7.8490
7.7696	6.7771		0.0425		0.69	7.6008
6.2554	6.5516		0.0816		0.69	7.3817
5.7343	6.7367		0.7524		0.69	7.5056
5.5367	6.2014		1.1300		0.69	7.0562
5.9926	6.1303		0.3447		0.69	6.8754
5.4649	5.9486		0.3842		0.69	6.6683
3.6099	8.8355		1.8607		0.91	9.0053
1.4080	8.6330		3.8193		0.91	9.4438
2.6532	6.2447		0.0265		0.91	9.0190
3.0278	7.0840		0.1631		0.91	8.7338
2.9118	7.7537		0.9542		0.91	8.4820
2.4459	7.9365		0.0493		0.91	8.6819
0.2341	1.9485		0.0622		1.01	1.9270
0.0028	1.0009		0.0491		1.01	1.8828

Figure 4.4 2^e itération du DGLM

Y	w	$\tilde{\mu}$	d	$\tilde{\gamma}$	$\tilde{\phi}$	w'
157.9510	16.5986	6.2405	4.2442	0.0281	1.03	10.3319
176.8613	16.2181	-0.0437	0.0373	-0.0969	1.03	10.0951
159.9696	15.4887	-0.0046	0.0121	-0.0322	1.03	9.6410
189.1498	14.9989	-0.0370	0.2013	-0.0316	1.03	9.3361
184.5313	14.5664	-0.0281	0.0638	-0.0283	1.03	9.0670
185.6152	14.9097	-0.0755	0.4480	-0.0204	1.03	9.2806
181.0256	13.9241	-0.0167	0.0761	-0.0241	1.03	8.6671
179.9608	13.6674	-0.0940	0.2488	-0.0267	1.03	8.4451
188.9111	13.1587	-0.0816	0.6960		1.03	8.1908
66.8696	22.6479	-1.1166	1.4339		0.93	46.0283
60.3142	22.3242	-3.2223	0.1940		0.93	44.9733
60.0330	21.3201	-4.2286	0.1840		0.93	42.9606
57.7061	20.6489	-4.6683	0.2928		0.93	41.5923
58.4448	20.0506	-5.4994	0.2202		0.93	40.3930
66.5860	20.6231	-5.8896	0.0461		0.93	41.3449
62.3546	19.1664	-5.9382	0.1671		0.93	38.6118
56.3601	18.6755	-6.6042	0.0673		0.93	37.6226
56.8690	18.1130	-6.8968	0.0960		0.93	36.4896
7.9273	8.4115	-11.1038	0.0696		1.00	9.1304
6.5306	8.2187		0.7877		1.00	8.9211
10.4381	7.8490		3.0273		1.00	8.5199
7.7696	7.6006		0.1093		1.00	8.2505
6.9564	7.3817		0.0637		1.00	8.0126
8.7343	7.5656		0.7636		1.00	8.2614
6.5387	7.0662		1.5097		1.00	7.6692
6.9326	6.6754		0.3252		1.00	7.4631
5.4649	6.6683		0.9573		1.00	7.2383
3.6099	9.6653		1.4395		1.00	10.2782
1.4080	9.4458		4.6177		1.00	10.0426
2.6632	9.0190		0.0155		1.00	9.5909
3.0278	8.7338		0.2846		1.00	9.2876
2.9118	8.4520		0.1106		1.00	9.0196
2.4499	8.6819		0.0236		1.00	9.2324
0.2341	1.9270		0.0724		1.00	1.8644
0.0028	1.8828		0.0671		1.00	1.8412

Figure 4.5 3^e itération du DGLM

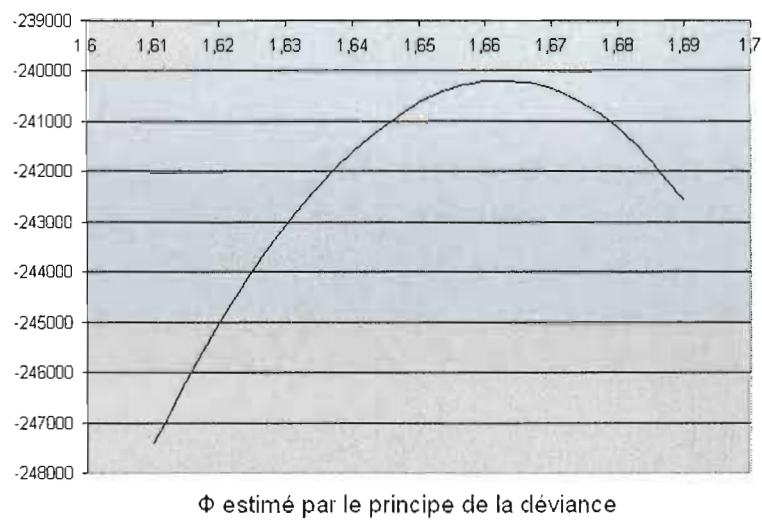


Figure 4.6 Optimisation simultanée de la fonction de vraisemblance 4.7 pour p , alors que les autres paramètres sont estimés par le principe de la déviance

CONCLUSION

En conclusion, les modèles stochastiques sont un outil indispensable pour décrire et capter la nature du risque des réserves actuarielles. Ils permettent d'associer une courbe de distribution aux pertes totales à venir, ce qui permet à la compagnie de mieux identifier le niveau de contingence des passifs et ainsi de pouvoir établir sa stratégie d'affaires en conséquence. Ces modèles tiennent compte de l'erreur statistique et d'estimation. Ils permettent aussi d'incorporer un jugement extérieur, et finalement, ils offrent la possibilité de tester la validité des paramètres utilisés.

Le mémoire présent a identifié les tendances contraires de la fréquence et la sévérité à mesure que le développement augmente et a proposé d'utiliser des modèles plus flexibles qui captent mieux la nature du risque. Les modèles de dispersion et les modèles linéaires généralisés doubles s'avèrent être des solutions naturelles qui ont déjà fait leur preuves dans d'autres domaines tel qu'en tarification.

N'empêche, peu importe la méthode utilisée, aussi perfectionnée soit-elle, on est sûr de ne pas prédire la perte exacte. C'est pourquoi il faut se rappeler que les méthodes statistiques ont pour but de prévoir l'avenir et non de le prédire.

BIBLIOGRAPHIE

- Actuarial Standards Board. 1991. « Actuarial Standard of Practice No. 9, Documentation and Disclosure in Property and Casualty Insurance Ratemaking, Loss Reserving, and Valuations ».
- Adler, M. et C. Kline Jr. 1978. « Evaluating Bodily Injury Liabilities Using a Claims Closure Model », *Evaluating Insurance Company Liabilities*, p. 1-66.
- Anderson, D., F. Feldblum, C. Modlin, D. Schirmacher, E. Schirmacher, et A. Thandi. 2004. « A Practitioner's Guide to Generalized Linear Models », *Casualty Actuarial Society Discussion Paper Program*.
- Bar-Lev, S. et P. Enis. 1984. *Reproducibility and Natural Exponential Families with Power Variance Functions*. Dept. of Statistics, State University of New York at Buffalo.
- Berquist, J. et R. Sherman. 1977. « Loss Reserve Adequacy Testing : A Comprehensive, Systematic Approach », *PCAS LXIV*.
- Brehm, P. 2002. « Correlation and the Aggregation of Unpaid Loss Distributions ». In *CAS Forum*, p. 1-23.
- Chib, S. et E. Greenberg. 1995. « Understanding the Metropolis-Hastings Algorithm », *American Statistician*, vol. 49, p. 327-335.
- Clark, D. 1996. « Basics of Reinsurance Pricing », *CAS Study Note*.
- Dennis, J. et H. Mei. 1979. « Two new unconstrained optimization algorithms which use function and gradient values », *Journal of Optimization Theory and Applications*, vol. 28, no. 4, p. 453-482.
- Dennis Jr, J., D. Gay, et R. Walsh. 1981. « An Adaptive Nonlinear Least-Squares Algorithm », *ACM Transactions on Mathematical Software (TOMS)*, vol. 7, no. 3, p. 348-368.
- Denuit, M. et A. Charpentier. 2004. *Mathématiques de l'assurance non-vie Tome I Principes fondamentaux de théorie du risque Économie et statistiques avancées*. Economica.
- Efron, B. et R. Tibshirani. 1993. « An Introduction to the Bootstrap », *Monographs on Statistics and Applied Probability*, vol. 57, p. 1-177.

- England, P. et R. Verrall. 1999. « Analytic and bootstrap estimates of prediction errors in claims reserving », *Insurance Mathematics and Economics*, vol. 25, no. 3, p. 281–293.
- England, P. et R. Verrall. 2002. « Stochastic Claims Reserving in General Insurance », *Institute of Actuaries and Faculty of Actuaries*.
- Fletcher, R. 1987. *Practical methods of optimization*. Wiley-Interscience New York, NY, USA.
- Gluck, S. et G. Venter. 2009. « Stochastic Trend Models in Casualty and Life Insurance ».
- Guszcza, J. et J. Lommele. 2006. « Loss Reserving Using Claim-Level Data ».
- Hachemeister, C. et J. Stanard. 1975. « IBNR Claims Count Estimation With Static Lag Functions ». In *ASTIN Colloquium, Portimao, Portugal and the 1976 CAS Spring Meeting Greenbrier*.
- Hougaard, P. 1986. « Survival models for heterogeneous populations derived from stable distributions », *Biometrika*, vol. 73, no. 2, p. 387–396.
- Jørgensen, B. 1997. *The Theory of Dispersion Models*. Chapman & Hall.
- Kaas, R. 2005. « Compound Poisson Distributions and GLM's - Tweedie's Distribution ».
- Kirschner, G., C. Kerley, et B. Isaacs. 2002. « Two Approaches to Calculating Correlated Reserve Indications Across Multiple Lines of Business ». In *CAS Forum Fall 2002*, p. 211–246.
- Klugman, S., H. Panjer, G. Willmot, et al. 1998. *Loss models : from data to decisions*. Wiley.
- Kunkler, M. 1998. « Modelling negatives in stochastic reserving models », *Insurance mathematics and economics*.
- Lee, Y. et J. Nelder. 2006. « Double hierarchical generalized linear models (with discussion) », *Journal of the Royal Statistical Society Series C*, vol. 55, no. 2, p. 139–185.
- Mack, T. 1991. « A simple parametric model for rating automobile insurance or estimating IBNR claims reserves », *ASTIN Bulletin*, vol. 22, no. 1, p. 93–109.
- . 1993. « Distribution-free calculation of the standard error of chain ladder reserve estimates », *ASTIN Bulletin*, vol. 23, no. 2, p. 213–225.
- . 1999. « The Standard Error of Chain Ladder Reserve Estimates : Recursive Calculation and Inclusion of a Tail Factor », *ASTIN Bulletin*, vol. 29, p. 361–366.

- Mildenhall, S. 1999. « A systematic relationship between minimum bias and generalized linear models ». In *Proceedings of the Casualty Actuarial Society*. T. 86, p. 93–487.
- Morris, C. 1981. Models for positive data with good convolution properties. Rapport, Rand Corporation. MEMO no. 8949.
- Nelder, J., Y. Lee, B. Bergman, A. Hynen, A. Huele, et J. Engel. 1998. « Joint Modeling of Mean and Dispersion », *Technometrics*, vol. 40, no. 2, p. 168–175.
- Nelder, J. et F. Wedderburn. 1989. *Generalized linear models*. Chapman and Hall.
- Olsson, D. et L. Nelson. 1975. « The Nelder-Mead simplex procedure for function minimization », *Technometrics*, vol. 17, no. 1, p. 45–51.
- Peters, G., P. Shevchenko, et M. Wüthrich. 2009. « Model uncertainty in claims reserving within Tweedie's compound Poisson models », *ASTIN Bulletin*, vol. 39, no. 1, p. 1–33.
- Renshaw, A. et R. Verrall. 1998. « A Stochastic Model Underlying the Chain-Ladder Technique », *British Actuarial Journal*, vol. 4, no. 4, p. 903–923.
- Ruoyan, M. 2004. « Estimation of Dispersion Parameters in GLMs with and without Random Effects ». Mémoire de maîtrise.
- Schiegl, M. 2009. « A three dimensional stochastic Model for Claim Reserving », *ASTIN Bulletin*.
- Smyth, G. 1989. « Generalized linear models with varying dispersion », *Journal of the Royal Statistical Society, Series B*, vol. 51, p. 47–60.
- Smyth, G., A. Huele, et A. Verbyla. 2001. « Exact and approximate REML for heteroscedastic regression », *Statistical Modelling*, vol. 1, no. 3, p. 161.
- Smyth, G. et B. Jorgensen. 2002. « Fitting Tweedie's Compound Poisson Model to Insurance Claims Data : Dispersion Modelling », *ASTIN Bulletin*, vol. 32, no. 1, p. 143–157.
- Taylor, G. 1987. « Regression models in claims analysis I : theory ». In *Proceedings of the Casualty Actuarial Society*. T. 74.
- Tweedie, M. 1984. « An index which distinguishes between some important exponential families ». In *Statistics : Applications and New Directions. Proceedings of the Indian Statistical Institute Golden Jubilee International Conference*, p. 579–604.
- Valdez, E., K. Antonio, et E. Frees. 2009. « A hierarchical model for micro-level stochastic loss reserving ».
- Verrall, R. 2000. « An investigation into stochastic claims reserving models and the chain-ladder technique », *Insurance Mathematics and Economics*, vol. 26, no. 1,

- p. 91–99.
- Verrall, R. et P. England. 2005. « Incorporating expert opinion into a stochastic model for the chain-ladder technique », *Insurance Mathematics and Economics*, vol. 37, no. 2, p. 355–370.
- Wiser, R. et A. Gardner. 1990. « Loss Reserving », *Foundations of Casualty Actuarial Science, Foundations of Casualty Actuarial Science, Casualty Actuarial Society*, p. 143–230.
- Wüthrich, M. 2003. « Claims Reserving Using Tweedie’s Compound Poisson Model », *ASTIN Bulletin*, vol. 33, no. 2, p. 331–346.
- Wüthrich, M. et M. Merz. 2008. *Stochastic Claims Reserving Methods in Insurance*. Wiley Finance.

CHAPITRE V

ANNEXES

5.1 Méthodes traditionnelles

5.1.1 London Chain

Année	1	2	3	4	5	6	7	8	9	10	11
1	17841110	25283543	26178956	26586700	26793830	26855399	26871377	26896301	26897537	26913180	26913501
2	19519117	26175637	27117095	27272490	27341948	27379717	27433549	27544940	27587203	27613036	27613365
3	19991172	26318655	27418832	27698481	27861135	27931135	27988013	27997894	28017550	28049741	28050076
4	19305646	25195437	25988457	26297499	26443420	26540885	26568408	26630328	26634450	26646206	26646523
5	18291478	24084760	24774204	25062830	25408354	25518939	25634782	25709137	25694898	25692771	25693078
6	18832520	24573734	25155532	25404095	25510970	25605182	25685856	25759490	25746254	25744887	25745194
7	17152710	23060996	23585802	23816258	24163162	24292955	24401845	24493617	24455149	24434705	24434997
8	16615059	21726236	22279513	22532390	22935159	23091182	23225913	23334295	23272720	23234805	23235082
9	16835453	21837350	22326706	22588091	22987586	23142489	23276116	23383789	23323201	23286032	23286310

Période	1-2	2-3	3-4	4-5	5-6	6-7	7-8	8-9	9-10	10-11
λ	1,2887	1,1112	1,0034	0,9412	0,9786	0,9785	0,9859	1,0199	1,0148	1,0000
α	712067	-1965758	185413	1727228	645901	631240	436462	-526675	-381775	0

Tableau 5.1 London Chain - Encouru cumulatif

5.1.2 Bornhuetter Ferguson

Année	1	2	3	4	5	6	7	8	9	10	11
1	17841110	25283543	26178956	26586700	26793830	26855399	26871377	26896301	26897537	26913180	26913501
2	19519117	26175637	27117095	27272490	27341948	27379717	27433549	27544940	27587203	27613036	27613364
3	19991172	26318655	27418832	27698481	27861135	27931135	27988013	27997894	28017550	28038848	28039182
4	19305646	25195437	25988457	26297499	26443420	26540885	26568408	26630328	26650980	26671518	26671840
5	18291478	24084760	24774204	25062830	25408354	25518939	25634782	25684316	25704204	25723981	25724291
6	18832520	24573734	25155532	25404095	25510970	25605182	25656207	25704789	25724294	25743691	25743995
7	17152710	23060996	23585802	23816258	24163162	24233591	24281615	24327339	24345697	24363952	24364238
8	16615059	21726236	22279513	22532390	22704445	22771939	22817962	22861781	22879373	22896868	22897143
9	16835453	21837350	22326706	22566954	22739008	22806503	22852526	22896345	22913937	22931432	22931707

Tableau 5.2 Bornhuetter Ferguson - Encouru cumulatif

5.2 Méthodes Stochastiques

5.2.1 Modèle de Poisson surdispersé

Année	1	2	3	4	5	6	7	8	9	10	11	Total
1	17841110	7442433	895413	407744	207130	61569	15978	24924	1236	15643	321	26913501
2	19519117	6656520	941458	155395	69458	37769	53832	111391	42263	25833		27613036
3	19991172	6327483	1100177	279649	162654	70000	56878	9881	19656			28017550
4	19305646	5889791	793020	309042	145921	97465	27523	61920				26630328
5	18291478	5793282	689444	288626	345524	110585	115843					25634782
6	18832520	5741214	581798	248563	106875	94212						25605182
7	17152710	5908286	524806	230456	346904							24163162
8	16615059	5111177	553277	252877								22532390
9	16835453	5001897	489356									22326706
Total	164384265	53872083	6568749	2172352	1384466	471600	270054	208116	63155	41476	321	229436637

Tableau 5.3 Encouru incrémental observé

Année	1	2	3	4	5	6	7	8	9	10	11	Total
1	19160589	6279317	765652	281127	201330	78979	53853	51275	20586	20472	321	26913500
2	19658846	6442606	785562	288437	206565	81033	55254	52608	21121	21004		27613037
3	19962020	6541962	797677	292885	209751	82282	56106	53420	21447			28017550
4	18988186	6222817	758763	278597	199518	78268	53369	50813				26630331
5	18313276	6001635	731793	268695	192427	75486	51472					25634784
6	18328929	6006765	732419	268924	192591	75551						25605179
7	17347880	5685255	693216	254530	182283							24163164
8	16300033	5341854	651345	239156								22532388
9	16324508	5349875	652323									22326705
Total	164384266	53872083	6568749	2172352	1384466	471600	270054	208116	63155	41476	321	229436639

Tableau 5.4 Modèle Poisson surdispersé - Encouru incrémental prévu

Année	1	2	3	4	5	6	7	8	9	10	11
1	17841110	25283543	26178956	26586700	26793830	26855399	26871377	26896301	26897537	26913180	26913501
2	19519117	26175637	27117095	27272490	27341948	27379717	27433549	27544940	27587203	27613036	27613357
3	19991172	26318655	27418832	27698481	27861135	27931135	27988013	27997894	28017550	28038861	28039187
4	19305646	25195437	25988457	26297499	26443420	26540885	26568408	26630328	26650738	26671010	26671319
5	18291478	24084760	24774204	25062830	25408354	25518939	25634782	25683779	25703464	25723015	25723314
6	18832520	24573734	25155532	25404095	25510970	25605182	25656705	25705745	25725447	25745015	25745314
7	17152710	23060996	23585802	23816258	24163162	24234669	24283432	24329845	24348492	24367012	24367295
8	16615059	21726236	22279513	22532390	22703669	22770861	22816681	22860294	22877815	22895217	22895483
9	16835453	21837350	22326706	22566228	22737760	22805050	22850938	22894615	22912162	22929590	22929856

Tableau 5.5 Modèle de Poisson surdispersé - Encouru cumulatif

5.2.2 Modèle de Poisson surdispersé - Courbe de Hoerl

Année	1	2	3	4	5	6	7	8	9	10	11
1	17841110	25283543	26178956	26586700	26793830	26855399	26871377	26896301	26897537	26913180	26913501
2	19519117	26175637	27117095	27272490	27341948	27379717	27433549	27544940	27587203	27613036	<i>28018690</i>
3	19991172	26318655	27418832	27698481	27861135	27931135	27988013	27997894	28017550	<i>26627333</i>	<i>26627398</i>
4	19305646	25195437	25988457	26297499	26443420	26540885	26568408	26630328	<i>25638966</i>	<i>25639183</i>	<i>25639245</i>
5	18291478	24084760	24774204	25062830	25408354	25518939	25634782	<i>25616893</i>	<i>25617650</i>	<i>25617867</i>	<i>25617929</i>
6	18832520	24573734	25155532	25404095	25510970	25605182	<i>24203835</i>	<i>24206331</i>	<i>24207045</i>	<i>24207250</i>	<i>24207309</i>
7	17152710	23060996	23585802	23816258	24163162	<i>22661215</i>	<i>22669388</i>	<i>22671725</i>	<i>22672395</i>	<i>22672587</i>	<i>22672642</i>
8	16615059	21726236	22279513	22532390	<i>22785946</i>	<i>22814766</i>	<i>22822994</i>	<i>22825348</i>	<i>22826022</i>	<i>22826215</i>	<i>22826270</i>
9	16835453	21837350	22326706	<i>26746907</i>	<i>26866205</i>	<i>26900185</i>	<i>26909887</i>	<i>26912662</i>	<i>26913457</i>	<i>26913684</i>	<i>26913750</i>

μ	α_2	α_3	α_4	α_5	α_6	α_7	α_8	α_9	γ_1	γ_2
18,0271	0,0257	0,0402	-0,0107	-0,0485	-0,0493	-0,1060	-0,1715	-0,1647	-0,0840	-1,2405

Tableau 5.6 Modèle de Poisson surdispersé avec courbe de Hoerl - Encouru cumulatif

5.2.3 Modèle gamma

Année	1	2	3	4	5	6	7	8	9	10	11
1	17841110	25283543	26178956	26586700	26793830	26855399	26871377	26896301	26897537	26913180	26913501
2	19519117	26175637	27117095	27272490	27341948	27379717	27433549	27544940	27587203	27613036	27613482
3	19991172	26318655	27418832	27698481	27861135	27931135	27988013	27997894	28017550	28037421	28037793
4	19305646	25195437	25988457	26297499	26443420	26540885	26568408	26630328	26649727	26670226	26670610
5	18291478	24084760	24774204	25062830	25408354	25518939	25634782	25702375	25728865	25756858	25757382
6	18832520	24573734	25155532	25404095	25510970	25605182	25647544	25691861	25709230	25727583	25727927
7	17152710	23060996	23585802	23816258	24163162	24243462	24292789	24344392	24364617	24385987	24386388
8	16615059	21726236	22279513	22532390	22690823	22755211	22794764	22836142	22852359	22869495	22869816
9	16835453	21837350	22326706	22544927	22693159	22753402	22790407	22829121	22844294	22860326	22860627

Tableau 5.7 Modèle gamma - Encouru cumulatif

5.2.4 Modèle Tweedie

Année	1	2	3	4	5	6	7	8	9	10	11
1	17841110	25283543	26178956	26586700	26793830	26855399	26871377	26896301	26897537	26913180	26913501
2	19519117	26175637	27117095	27272490	27341948	27379717	27433549	27544940	27587203	27613036	<i>27613362</i>
3	19991172	26318655	27418832	27698481	27861135	27931135	27988013	27997894	28017550	<i>28038783</i>	<i>28039115</i>
4	19305646	25195437	25988457	26297499	26443420	26540885	26568408	26630328	<i>26650588</i>	<i>26670730</i>	<i>26671044</i>
5	18291478	24084760	24774204	25062830	25408354	25518939	25634782	<i>25684293</i>	<i>25704091</i>	<i>25723773</i>	<i>25724080</i>
6	18832520	24573734	25155532	25404095	25510970	25605182	<i>25655929</i>	<i>25704492</i>	<i>25723911</i>	<i>25743216</i>	<i>25743517</i>
7	17152710	23060996	23585802	23816258	24163162	<i>24234770</i>	<i>24283432</i>	<i>24330001</i>	<i>24348623</i>	<i>24367135</i>	<i>24367424</i>
8	16615059	21726236	22279513	22532390	<i>22702489</i>	<i>22769232</i>	<i>22814588</i>	<i>22857994</i>	<i>22875350</i>	<i>22892605</i>	<i>22892874</i>
9	16835453	21837350	22326706	<i>22564116</i>	<i>22733819</i>	<i>22800407</i>	<i>22845658</i>	<i>22888963</i>	<i>22906279</i>	<i>22923493</i>	<i>22923762</i>

Tableau 5.8 Modèle Tweedie avec expositions, $p = 1.1741$ - Encouru cumulatif

5.3 Modèles de dispersion

5.3.1 Modèle Tweedie

Année	1	2	3	4	5	6	7	8	9	10	11
1	17841110	25283543	26178956	26586700	26793830	26855399	26871377	26896301	26897537	26913180	26913501
2	19519117	26175637	27117095	27272490	27341948	27379717	27433549	27544940	27587203	27613036	<i>27613360</i>
3	19991172	26318655	27418832	27698481	27861135	27931135	27988013	27997894	28017550	<i>28038574</i>	<i>28038902</i>
4	19305646	25195437	25988457	26297499	26443420	26540885	26568408	26630328	<i>26650317</i>	<i>26670202</i>	<i>26670512</i>
5	18291478	24084760	24774204	25062830	25408354	25518939	25634782	<i>25683373</i>	<i>25702590</i>	<i>25721708</i>	<i>25722006</i>
6	18832520	24573734	25155532	25404095	25510970	25605182	<i>25655929</i>	<i>25704649</i>	<i>25723917</i>	<i>25743085</i>	<i>25743385</i>
7	17152710	23060996	23585802	23816258	24163162	<i>24234260</i>	<i>24282498</i>	<i>24328810</i>	<i>24347125</i>	<i>24365346</i>	<i>24365631</i>
8	16615059	21726236	22279513	22532390	<i>22702178</i>	<i>22768672</i>	<i>22813787</i>	<i>22857100</i>	<i>22874229</i>	<i>22891270</i>	<i>22891536</i>
9	16835453	21837350	22326706	<i>22564283</i>	<i>22733784</i>	<i>22800167</i>	<i>22845205</i>	<i>22888444</i>	<i>22905545</i>	<i>22922558</i>	<i>22922823</i>

Tableau 5.9 Modèle Tweedie avec expositions, $p = 1.8111$ - Encouru cumulatif

Paramètre	Estimé	Écart-type	Prob(> T)
μ	5,1540	0,0236	0,0001
α_2	0,0334	0,0322	0,3055
α_3	0,0913	0,0325	0,0074
α_4	0,0677	0,0328	0,0449
α_5	0,0576	0,0331	0,0885
α_6	0,0370	0,0329	0,2674
α_7	0,0547	0,0335	0,1099
α_8	0,0137	0,0339	0,6882
α_9	0,0426	0,0342	0,2202
β_2	-1,1144	0,0175	0,0001
β_3	-3,2208	0,0438	0,0001
β_4	-4,2209	0,0833	0,0001
β_5	-4,5585	0,1377	0,0001
β_6	-5,4959	0,2069	0,0001
β_7	-5,8838	0,3289	0,0001
β_8	-5,9246	0,4202	0,0001
β_9	-6,8522	0,5942	0,0001
β_{10}	-6,8574	0,8068	0,0001
β_{11}	-11,0173	1,6832	0,0001
ϕ	5,4798	0,0039	0,0001
ϕ_2	0,5302	0,0069	0,0001
ϕ_3	2,3013	0,0171	0,0001
ϕ_4	3,3332	0,0326	0,0001
ϕ_5	4,1649	0,0539	0,0001
ϕ_6	4,6658	0,0810	0,0001
ϕ_7	5,3461	0,1287	0,0001
ϕ_8	5,6216	0,1645	0,0001
ϕ_9	5,8678	0,2326	0,0001
ϕ_{10}	6,0879	0,2848	0,0001

Tableau 5.10 Modèle Tweedie avec expositions, $p = 1.8111$ - paramètres

5.4 Modèles linéaires généralisés doubles

5.4.1 Modèle Tweedie

Année	1	2	3	4	5	6	7	8	9	10	11
1	17841110	25283543	26178956	26586700	26793830	26855399	26871377	26896301	26897537	26913180	26913501
2	19519117	26175637	27117095	27272490	27341948	27379717	27433549	27544940	27587203	27613036	<i>27613036</i>
3	19991172	26318655	27418832	27698481	27861135	27931135	27988013	27997894	28017550	<i>28017550</i>	<i>28017550</i>
4	19305646	25195437	25988457	26297499	26443420	26540885	26568408	26630328	<i>26630328</i>	<i>26630328</i>	<i>26630328</i>
5	18291478	24084760	24774204	25062830	25408354	25518939	25634782	<i>25634782</i>	<i>25634782</i>	<i>25634782</i>	<i>25634782</i>
6	18832520	24573734	25155532	25404095	25510970	25605182	<i>25605182</i>	<i>25605182</i>	<i>25605182</i>	<i>25605182</i>	<i>25605182</i>
7	17152710	23060996	23585802	23816258	24163162	<i>24163162</i>	<i>24163162</i>	<i>24163162</i>	<i>24163162</i>	<i>24163162</i>	<i>24163162</i>
8	16615059	21726236	22279513	22532390	<i>22532390</i>	<i>22532390</i>	<i>22532390</i>	<i>22532390</i>	<i>22532390</i>	<i>22532390</i>	<i>22532390</i>
9	16835453	21837350	22326706	<i>22326706</i>	<i>22326706</i>	<i>22326706</i>	<i>22326706</i>	<i>22326706</i>	<i>22326706</i>	<i>22326706</i>	<i>22326706</i>

Tableau 5.11 Modèle Tweedie avec expositions, $p = 1.6625$ - Encouru cumulatif

Paramètre	Estimé
μ	16,9187
α_2	-0,0977
α_3	-0,1262
α_4	-0,1956
α_5	-0,2134
α_6	-0,2307
α_7	-0,2172
α_8	-0,3448
α_9	-0,3660
β_2	-1,1213
β_3	-3,2330
β_4	-4,2330
β_5	-4,5609
β_6	-5,4891
β_7	-5,8742
β_8	-5,9461
β_9	-6,8696
β_{10}	-6,9219
β_{11}	-11,1473

Période	1	2	3	4	5	6
Exposition w_j	0,3708	4,2721	0,2351	0,2310	0,0459	0,1170

7	8	9	10	11
0,0422	0,0280	0,0280	0,0280	0,0280

Tableau 5.12 Modèle Tweedie avec expositions, $p = 1.6625$ - paramètres